



```
FT      /*tag- a
PN      /label- I-3-P_binding-peptide
PD      J03183482-A.
PE      09-AUG-1991.
PR      14-DEC-1989; 324256.
PA      (KYOW ) KYOWA HAKKO KOGYO KK.
DR      WPI; 91-277584/38.
DR      P-PDB; R1387.
PT      New polypeptide having binding affinity to inositol-3-phosphate -
PT      prep. by culturing cell contg. recombinant plasmid comprising
PT      DNA and vector DNA
PS      Disclosure: Fig 2(1-3); 11pp; Japanese.
CC      The sequence may be included in a plasmid/vector for transformation
CC      of a host cell and mass-prodn. of the peptide.
SQ      Sequence 9871 Bp; 2687 A; 2325 C; 2555 G; 2304 T;

Query Match
Best Local Similarity 100.0%; Pred. No. 22;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      93 CCCTCCAGACAGCA 108
Db      2432 CCCTCCAGACAGCA 2417

RESULT  3
Q25975
ID      025975 standard: DNA; 15377 Bp.
AC      025975;
DT      08-JAN-1993 (first entry)
DE      MH mutant porcine ryanodine receptor cDNA.
KM      MH; RYR1; calcium release channel; sarcoplasmic reticulum;
KW      transverse tubule; Plectrain; Yorkshire; polymorphism; beta strand; ss.
OS      Synthetic.
FH      Key
FH      cds
FT      130..15237
FT      Location/Qualifiers
FT      /*tag= a
FT      207
FT      variation
FT      /*tag= b
FT      405
FT      variation
FT      /*tag= c
FT      438
FT      variation
FT      /*tag= d
FT      876
FT      variation
FT      /*tag= e
FT      1329
FT      variation
FT      /*tag= f
FT      1972
FT      variation
FT      /*tag= g
FT      2007
FT      variation
FT      /*tag= h
FT      4071
FT      variation
FT      /*tag= i
FT      4383
FT      variation
FT      /*tag= j
FT      4462
FT      variation
FT      /*tag= k
FT      4494
FT      variation
FT      /*tag= l
FT      6867
FT      variation
FT      /*tag= Polymorphic_site
```

```
FT      /*tag= m
FT      /label= Polymorphic_site
FT      7692
FT      variation
FT      /*tag= n
FT      /label= Polymorphic_site
FT      8940
FT      variation
FT      /*tag= o
FT      9192
FT      variation
FT      /*tag= p
FT      /label= Polymorphic_site
FT      9585
FT      variation
FT      /*tag= q
FT      /label= Polymorphic_site
FT      9600
FT      variation
FT      /*tag= r
FT      /label= Polymorphic_site
FT      9951
FT      variation
FT      /*tag= s
FT      /label= Polymorphic_site
FT      10111
FT      variation
FT      /*tag= t
FT      /label= Polymorphic_site
FT      11250
FT      variation
FT      /*tag= u
FT      /label= Polymorphic_site
FT      12300
FT      variation
FT      /*tag= v
FT      /label= Polymorphic_site
FT      14007
FT      variation
FT      /*tag= w
FT      /label= Polymorphic_site
FT      15355..15360
FT      /*tag= x

polya_signal
FT      15355..15360
FT      /*tag= x

PN      W09211387-A.
PD      09-JUL-1992.
PE      20-DEC-1991; CA0457.
PR      21-DEC-1990; GB-027869.
PR      20-MAY-1991; GB-010865.
PR      09-SEP-1991; GB-019250.
PA      (UYGU-) UNIV GUELPH.
PA      (UTOR ) UNIV TORONTO INNOVATIONS FOUND.
PI      MacLennan DH, O'Brien PJ;
DR      WPI; 92-250106/30.
DR      P-PDB; R25450.
PT      Purified DNA mol. for diagnosis of porcine malignant hyperthermia
PT      - comprises DNA sequence encoding normal or mutant ryanodine
PT      receptor with specified endonuclease restriction map
PS      Disclosure: Fig 2; 96pp; English.
CC      The sequence given is the mutant pig ryanodine receptor (RYR1) gene
CC      from swine cDNA. The polymorphic sites were observed in comparisons
CC      of Plectrain and Yorkshire breeds. There are 17 polymorphisms between
CC      the two breeds. The polymorphism at position 1972 causes a mutation
CC      from Arg to Cys and this is thought to be the molecular basis of
CC      porcine malignant hyperthermia (MH). This mutation lies within the
CC      region of RYR1 that is concerned with the binding of regulators of Ca2+
CC      release channel gating. Analysis of surrounding sequences suggests
CC      that this mutation lies within a beta strand domain comprising roughly
CC      of amino acids 520 to 830. RYR1 is the calcium release channel of the
CC      sarcoplasmic reticulum and is a large protein which spans the gap
CC      between the transverse tubule and the sarcoplasmic reticulum. The
CC      channel is activated by ATP, calcium, caffeine, and micro-molar
CC      ryanodine. It is inhibited by ruthenium red, tetracaine, calmodulin,
CC      high Mg2+ and ryanodine.
SQ      Sequence 15377 Bp; 3197 A; 4630 C; 4755 G; 2774 T;
```

Query Match  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 TCATTGATGCTTTGG 27  
|||||



Db 14930 TCATTGATGCTTTGG 14945

# RESULT 4

OS1025  
ID 051025 standard: cDNA: 2718 BP.  
AC 051025:  
DT 18-APR-1994 (first entry)  
DE Human glutamate receptor gene, HSGIUR1.  
KW L-glutamate; receptors; neurological disorders; stroke; epilepsy;  
KM head trauma; neurodegenerative processes; Alzheimer's disease; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 1..2718  
FT FT /\*tag- a  
FT misc-feature 1..60  
FT FT /\*tag- b  
FT misc-feature 130..189  
FT FT /\*tag- c  
FT FT /\*tag- "probe sequence"  
FT FT 136..185  
FT FT /\*tag- d  
FT FT /\*tag- "probe sequence"  
FT misc-feature 2662..2718  
FT FT /\*tag- e  
FT FT /\*tag- "probe sequence with TAA added at the 3' end"  
FT EP-568384-A.  
PN 03-NOV-1993.  
PF 30-APR-1993: 303401.  
PR 01-MAY-1992: US-879688.  
PR 19-MAY-1992: US-885912.  
PA (ELIL) Lilly & CO Ltd.  
PI Burnett JP, Mayne NG, Sharp RL, Snyder YM;  
DR WPI: 93-346751/44.  
DR P-PSDB: R42354.  
PT Human glutamate receptor protein and nucleic acid - used to  
PT develop therapeutic agents for neurological disorders and  
PT neuro-degenerative diseases  
PS Claim 5: Page 27-31: 42pp: English.  
CC The DNA sequence encodes the human glutamate receptor HSGIUR1. The  
CC cDNA was produced from a culture of E. coli cells contg. the plasmid  
CC pS103 (obtd. from American Type Culture Collection, Rockville  
CC Maryland). The plasmid was digested with restriction enzymes EcoRI  
CC and KpnI to produce a 4.2 kb fragment contg. the entire coding  
CC sequence of the human HSGIUR1 receptor. The receptor can be used to  
CC study interactions of L-glutamate with its receptor and to develop  
CC therapeutic agents for neurological disorders such as stroke,  
CC epilepsy and head trauma and neurodegenerative disorders such as  
CC Alzheimer's disease. Fragments of the DNA sequence may be used as  
CC probes to identify further glutamate receptor genes.  
CC See also 051026-32.  
SO Sequence 2718 BP: 703 A: 636 C: 740 G: 639 T:

Query Match 7.38: Score 16; DB 1; Length 2718;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 ATGAATGCACAGTGA 128  
DB 1688 ATGAATGCACAGTGA 1703

# RESULT 5

OS4116  
ID 054116 standard: cDNA: 3220 BP.  
AC 054116:  
DT 29-JUN-1994 (first entry)  
DE Human GluR1B receptor DNA.  
KW GluR1B receptor; glutamate receptor; AMPA-binding receptor;  
KM CNS: central nervous system; plasmid pBS/humGluR1B; ds.  
OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 62..2782  
FT signal\_peptide 62..115  
FT FT /\*tag- a  
FT FT /\*tag- b  
FT mat\_peptide 116..2782  
FT FT /\*tag- c  
FN EP-574257-A.  
PD 15-DEC-1993.  
PF 10-JUN-1993: 304500.  
PR 10-JUN-1992: US-896437.  
PR 10-JUN-1992: US-896611.  
PR 10-JUN-1992: US-896612.  
PA (ALX) ALLEX BIOFARMACEUTICALS.  
PA (ELIL) ELIOTT C.  
PA (KAMBOJ) R.  
PA KAMBOJ R.  
PA (NUTT) NUTT S. L.  
PI Elliott C, Kamboj R, Nutt SL;  
DR WPI: 93-396903/50.  
DR P-PSDB: R44300.  
PT Polynucleotide(s) encoding AMPA-binding human glutamate receptors  
PT - used for providing prods. for use in screening assays or in  
PT detection and/or localisation  
PS Claim 1: Page 15-19: 10pp: English.  
CC The GluR1B receptor DNA is contained within plasmid pBS/humGluR1B  
CC (ATCC 75246) and may be expressed by a mammal recombinant cell. The  
CC cDNA can be used to obtain products for evaluating the human nervous  
CC system and particularly for assessing potentially therapeutic  
CC interactions between the AMPA-binding human excitatory amino acid  
CC receptors and natural and synthetic ligands. It can also be used  
CC for detecting the presence and/or location of human GluR receptors  
CC e.g. in brain tissue. The DNA was isolated from human brain cDNA  
CC libraries using rat GluR receptor gene probes.  
SO Sequence 3220 BP: 844 A: 751 C: 837 G: 788 T:

Query Match 7.38: Score 16; DB 1; Length 3220;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 ATGAATGCACAGTGA 128  
DB 1749 ATGAATGCACAGTGA 1764

# RESULT 6

OS6656/C  
ID 06656 standard: cDNA: 2216 BP.  
AC 06656:  
DT 19-JAN-1995 (first entry)  
DE Murine delta opioid receptor coding sequence.  
KW delta opioid; enkephalin; receptor; mouse; murine; analgesic; pain;  
KM drug addiction; neurological disorder; psychiatric; disorder;  
KW cardiovascular disorder; ds.  
OS Mus musculus.  
FH Key Location/Qualifiers  
FT cds 59..1174  
FT FT /\*tag- a  
FT FT /\*tag- a  
FT FT /product- opioid\_receptor  
FN ER2697850-A.  
PD 13-MAY-1994.  
PF 10-NOV-1992: 013526.  
PR 10-NOV-1992: FR-013526.  
PA (UYST-) UNIV PASTEUR STRASBOURG LODIS.  
PI Kieffer B;  
DR WPI: 94-178255/22.  
DR P-PSDB: R66503.  
PT New nucleic acid encoding opioid receptor - and related  
PT polypeptide, antisense nucleic acid, probes, recombinant cells  
PT and ligands, useful in diagnosis and treatment of e.g.  
PT neurological disorders  
PS Claim 3: Page 16-18: 29pp: French.

CC A cDNA bank constructed from hybridoma NG108-15, was used to  
 CC transfect COS-1 cells. The cells were tested for ability to bind  
 CC tritium-labelled Tyr-D-Thr-Phe-Leu-Thr, in the presence or  
 CC absence of the oploid antagonist naloxone. Clone K56 was isolated  
 CC from a positive colony and found to contain a 2216bp insert. This  
 CC cDNA encodes a delta oploid (enkephalin) receptor with apparent  
 CC dissociation constant 1.4nM and Bmax 3.9-6.4 pmole/mg protein.  
 SQ Sequence 2216 BP; 460 A; 647 C; 649 G; 460 T;

Query Match 7.3%; Score 16; DB 1; Length 2216;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 TGCTTTGGGGCCCTT 34  
 DB 2131 TGCTTTGGGGCCCTT 2116

RESULT 7  
 Q75927/c  
 ID Q75927 standard; DNA; 2272 BP.  
 AC Q75927;  
 DT 17-AUG-1995 (first entry)  
 DE Mouse delta oploid receptor MOR1 cDNA.  
 KM Mouse; kappa; delta; mu; oploid receptor; brain; primer; PCR; amplify;  
 KW transmembrane domain; somatostatin; receptor; human; expression vector;  
 KW truncate; chimeric; assay; probe; ss.  
 OS Mus musculus.  
 FH Key Location/Qualifiers  
 FT cds 12..1130  
 FT /tag= a  
 FT /product= mouse delta oploid receptor

PN W09428132-A.  
 PD 08-DEC-1994.  
 PF 20-MAY-1994; U05747.  
 PR 20-MAY-1993; US-066296.  
 PR 30-JUL-1993; US-100694.  
 PR 05-NOV-1993; US-147592.  
 PA (ARCH-) ARCH DEV CORP.  
 PI Bell GJ, Reisine T, Yasuda K;  
 DR WPI: 95-022804/03.  
 DR P-PDB: R67670.  
 PT Polynucleotides and peptides derived from oploid receptor  
 PT polypeptides for use in therapeutic compositions and in  
 PT screening assays for useful drug substances.  
 PS Claim 6; Page 215-221; 300pp; English.  
 CC The nucleotide sequence of the novel mouse delta oploid receptor gene  
 CC MOR1. The gene was isolated from a mouse brain cDNA library using a  
 CC fragment (amplified from the cDNA library with primers 075929-30) as a  
 CC probe. The primers are based on the conserved sequences present in the  
 CC second and third transmembrane domains of somatostatin (SRII) receptor  
 CC subtypes SSTR1, SSTR2 and SSTR3. The 1.3 kb EcoRI-SacI fragment from the  
 CC mouse delta oploid receptor clone, lambda ms1-2, was subcloned into the  
 CC CMV promoter-based expression vector PCMV-6c. The resultant construct  
 CC PCMV-ms1-2 was transfected into COS-1 cells for protein production. The  
 CC gene encoding the oploid receptor can be used to produce complete,  
 CC truncated or chimeric oploid receptor proteins. The oploid receptors  
 CC thus produced are useful for the development of novel assays designed to  
 CC select or improve substances, capable of interacting with the oploid  
 CC receptor proteins, for use in diagnosis, drug design and therapeutic  
 CC applications.  
 SQ Sequence 2272 BP; 485 A; 665 C; 650 G; 472 T;

Query Match 7.3%; Score 16; DB 1; Length 2272;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 TGCTTTGGGGCCCTT 34  
 DB 2073 TGCTTTGGGGCCCTT 2058

RESULT 8  
 T02795  
 ID T02795 standard; cDNA to mRNA; 2911 BP.  
 AC T02795;  
 DT 24-APR-1996 (first entry)  
 DE Human glutamate receptor subunit cDNA (gluRAflpG).  
 KW glutamate receptor subunit; human; flip; flop; splice variant;  
 KW identify ligand; treatment; ischemia; schizophrenia; epilepsy; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 144..2864  
 FT /tag= a  
 FT /product= glutamate receptor subunit

PN W09521188-A1.  
 PD 10-AUG-1995.  
 PF 27-JAN-1995; E00290.  
 PR 07-FEB-1994; DE-403666.  
 PA (BADI) BASF AG.  
 PI Bach A, Hoeger T, Lemaire H, Sterrer S, Ullsch A;  
 DR WPI: 95-283726/37.  
 DR P-PDB: R84912.  
 PT DNA sequences coding for glutamate receptor sub-unit variants  
 PT useful for the identification of functional ligands for these  
 PT receptors and for preparing anti-sense reagents  
 PS Claim 1; Page 13-18; 74pp; English.  
 CC T02795-802 are DNA sequences encoding variants of the human glutamate  
 CC receptor (gluR) subunits. cDNAs encoding gluR subunits A, B, C, and D,  
 CC were isolated from the brain. There are 2 splice variants, "flip" and  
 CC "flop", for each subunit that are known. It was found that for gluRA,  
 CC gluRB, gluRC and gluRD, the first amino acid of the "flip/flop"-region,  
 CC can be Glycine (G) or Arginine (R), through different RNA splicing.  
 CC For the gluRB subunit, it was found that it is possible to have a  
 CC Glutamine (Q) or an R in the transmembrane domain, before the flip/flop  
 CC region giving a total of 8 possible conformations. The gluRAflpR cDNA  
 CC possesses in comparison to gluRAflpG cDNA, a base change at position  
 CC 2269, a glycine codon (GGA) is converted to an arginine codon (AGA)  
 CC (sic). The same applies for gluRAflopR. Cells transfected with the DNA  
 CC are useful for identifying functional ligands for gluR. Glutamate is  
 CC an important excitatory neurotransmitter in the central nervous system  
 CC and is involved in the pathophysiology of e.g., epilepsy, schizophrenia  
 CC and ischemia. gluR ligands have the potential to be used in the  
 CC treatment of these conditions.  
 SQ Sequence 2911 BP; 775 A; 674 C; 799 G; 663 T;

Query Match 7.3%; Score 16; DB 1; Length 2911;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 ATGATGCGACAGTGA 128  
 DB 1831 ATGATGCGACAGTGA 1846

RESULT 9  
 T02796  
 ID T02796 standard; cDNA to mRNA; 2946 BP.  
 AC T02796;  
 DT 26-APR-1996 (first entry)  
 DE Human glutamate receptor subunit cDNA (gluRAflpG).  
 KW glutamate receptor subunit; human; flip; flop; splice variant;  
 KW identify ligand; treatment; ischemia; schizophrenia; epilepsy; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 144..2864  
 FT /tag= a  
 FT /product= glutamate receptor subunit  
 PN W09521188-A1.  
 PD 10-AUG-1995.  
 PF 27-JAN-1995; E00290.  
 PR 07-FEB-1994; DE-403666.  
 PA (BADI) BASF AG.



Key	Location/Qualifiers	Score	DB	Length	Indels	Gaps
FT	50..1204					
FT	/*tag= a					
PD	MO9843478-A1.					
PF	08-OCT-1998.					
PR	01-APR-1998; U06371.					
PR	29-JUL-1997; US-902615.					
PR	01-APR-1997; US-833457.					
PR	24-JUN-1997; US-881227.					
PA	(HUMA-) HUMAN GENOME SCI INC.					
PA	(INMR-) MERIEUX ORAYAV PASTEUR MERIEUX SERUMS.					
PI	Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;					
DR	WPI: 98-542293/46.					
DR	P-PSDB; W98586.					
PT	New isolated Helicobacter polynucleotides - used to develop products					
PT	for the diagnosis, prevention and treatment of Helicobacter					
PT	infections and gastrointestinal diseases					
PS	Claim 1; Page 1159-1161; 2054pp; English.					
CC	This sequence represents a polynucleotide of the invention. It was					
CC	isolated from Helicobacter pylori and encodes a H.pylori GHP protein.					
CC	The polypeptides can be used for preventing or treating Helicobacter					
CC	infections, and gastroduodenal diseases associated with these					
CC	infections, including acute, chronic, and atrophic gastritis, and peptic					
CC	ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used					
CC	for the production of antibodies. The products can also be used for					
CC	detection and diagnosis:					
SQ	Sequence 1304 BP; 372 A; 258 C; 324 G; 350 T;					
Query Match	7.3%; Score 16; DB 1; Length 1304;					
Best Local Similarity	100.0%; Pred. No. 22;					
Matches 16; Conservative	0; Mismatches 0; Indels 0; Gaps 0					
QY	78 AGAGTAGGCGATTAC 93					
DB	1198 AGAGTAGGCGATTAC 1213					
RESULT 13						
X14113						
ID	X14113 standard; DNA: 1837 BP.					
AC	X14113;					
DT	31-MAR-1999 (first entry)					
DE	H. pylori GHP 84 gene.					
KW	GHP protein; Helicobacter infection; gastroduodenal disease; gastritis;					
KW	peptic ulcer disease; ss.					
OS	Helicobacter pylori.					
FT	Key					
FT	Location/Qualifiers					
FT	51..1787					
FT	/*tag= a					
PD	MO9843478-A1.					
PF	08-OCT-1998.					
PF	01-APR-1998; U06371.					
PR	29-JUL-1997; US-902615.					
PR	01-APR-1997; US-833457.					
PR	24-JUN-1997; US-881227.					
PA	(HUMA-) HUMAN GENOME SCI INC.					
PA	(INMR-) MERIEUX ORAYAV PASTEUR MERIEUX SERUMS.					
PI	Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;					
DR	WPI: 98-542293/46.					
DR	P-PSDB; W98394.					
PT	New isolated Helicobacter polynucleotides - used to develop products					
PT	for the diagnosis, prevention and treatment of Helicobacter					
PT	infections and gastrointestinal diseases					
PS	Claim 1; Page 641-644; 2054pp; English.					
CC	This sequence represents a polynucleotide of the invention. It was					
CC	isolated from Helicobacter pylori and encodes a H.pylori GHP protein.					
CC	The polypeptides can be used for preventing or treating Helicobacter					
CC	infections, and gastroduodenal diseases associated with these					
CC	infections, including acute, chronic, and atrophic gastritis, and peptic					
CC	ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used					
CC	for the production of antibodies. The products can also be used for					
CC	detection and diagnosis.					

```

SQ      Sequence 1837 BP; 552 A; 347 C; 439 G; 499 T;

Query Match
Best Local Similarity 100.0%; Score 16; DB 1; Length 1837;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      78 ACAGTAAGGATTAAAC 93
        |||||||
DB      35 ACAGTAAGGATTAAAC 50

RESULT 14
ID      004525/c
AC      004525:
DT      01-OCT-1990 (first entry)
DE      Total base sequence of rice plant chloroplast DNA.
KW      Chloroplast; true grass; rice plant; ss.
OS      Oryza sativa.
PN      J02100682-A.
PD      12-APR-1990.
PF      07-OCT-1988; 251967.
PR      07-OCT-1988; JP-251967.
PA      WPII; 90-159709/21.
PT      (MITR) Mitsui Toatsu Chem Inc.
PR      Chloroplast DNA of true grasses - used to produce various
PS      DNA base sequences by decomposition of rice plant DNA.
CS      Claim 1; Fig 1; 20pp; Japanese.
SQ      The sequence is that of the whole of rice chloroplast DNA.
        Sequence 134525 BP; 41249 A; 26129 C; 26331 G; 40816 T;

Query Match
Best Local Similarity 100.0%; Score 15; DB 1; Length 134525;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      198 TTCCTCTGGGATCCCA 212
        |||||||
DB      26807 TTCCTCTGGGATCCCA 26793

RESULT 15
ID      051227/c
AC      051227:
DT      22-MAY-1994 (first entry)
DE      Human MSH-R gene.
KW      Melanocyte stimulating hormone receptor; mouse; genetic disease:
        Isolation; screening; antagonists; ss.
OS      Homo sapiens.
FH      Key
FH      Location/Qualifiers
FT      cds
FT      462..1415
FT      /*tag= a

MO9321316-A.
PD      28-OCT-1993.
PF      07-APR-1993; U03247.
PR      10-APR-1992; US-866979.
PA      (UOR-) UNIV OREGON HEALTH SCI.
PI      Cone BD, MountJoy KG;
DR      WPI; 93-351726/44.
P-PSDB: R43571.

Melanocyte stimulating hormone receptors and corresp. DNA - from
humans and mice, used for diagnosis of genetic disease and for
screening agents
Claim 5; Fig 1B; 46pp; English.

The sequence encodes the human melanocyte-stimulating hormone
receptor isolated from a human melanoma cDNA library using a pair
of degenerate PCR primers corresp. to the putative third and sixth
transmembrane regions of G-protein coupled receptors. The sequence
can be used as a probe for the detection and diagnosis of genetic
diseases, or for the isolation of novel mammalian receptors and genes.
It may also be used for screening and quantitating cnds. as inhibitors

```

Mon Aug 9 13:27:14 1999

us-09-049-696-9\_1.rng

Page 7

CC of agonist binding to the receptor.  
CC See also Q51226 and Q57564-5.  
SQ Sequence 1633 BP; 292 A; 519 C; 497 G; 325 T;

Query Match 6.9%; Score 15; DB 1; Length 1633;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 CTCTCTCGGATCC 211  
|||||  
Db 487 CTCTCTCGGATCC 473

Search completed: August 5, 1999, 17:40:43  
Job time: 6278 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:12:11 ; Search time 3264.21 Seconds  
(without alignments)  
212.397 Million cell updates/sec

Title: US-09-049-696-9

Perfect score: 218  
Sequence: 1 GAACAATGGCCCTCATTTGATG.....TCTCTGGATGCCAGTGCAC 218

Scoring table: IDENTITY-MUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBank: \*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hcg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_id:\*  
23: em\_cm:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hcg1:\*  
35: gb\_hcg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	218	100.0	3311	11	AF039400	Homo sapi

2	218	100.0	2826	42	AF127036	AF127036 Homo sapi
3	156.2	71.7	2937	12	AB017156	AB017156 Mus muscu
4	148	67.9	35278	11	AF039401	AF039401 Homo sapi
5	89.8	41.2	2984	3	BT036445	BT036445 Bos taurus
6	88.2	40.5	3317	3	AF001261	AF001261 Bos tauru
7	88.2	40.5	3288	3	AF001262	AF001262 Bos tauru
8	88.2	40.5	2820	3	AF001263	AF001263 Bos tauru
9	88.2	40.5	3415	42	AF043976	AF043976 Homo sapi
10	80.2	36.8	3022	12	AF047838	AF047838 Mus muscu
11	80.2	36.8	3471	12	AF052746	AF052746 Mus muscu
12	72.8	33.4	3604	9	AB026833	AB026833 Homo sapi
13	32.6	15.0	2414	37	DMU46886	DMU46886 Drosophila
14	32.2	14.8	3197	12	AF041431	AF041431 Mus muscu
15	31.8	14.6	201175	34	AC004676	AC004676 Homo sapi
16	31.6	14.5	152927	42	AC004895	AC004895 Homo sapi
17	31	14.2	122325	9	HS1044017	HS1044017 Human DNA
18	31	14.2	7876	17	SVBVCOMGN	SVBVCOMGN
19	30.8	14.1	130755	11	AC004253	AC004253 Homo sapi
20	30.8	14.1	166909	34	HS105613	HS105613 Homo sapi
21	30.6	14.0	100000	9	AP000052	AP000052 Homo sapi
22	30.6	14.0	100000	9	AP000121	AP000121 Homo sapi
23	30.6	14.0	100000	9	AP000167	AP000167 Homo sapi
24	30.6	14.0	16145	9	AP000321	AP000321 Homo sapi
25	30.4	13.9	212929	34	HS21242	HS21242 Homo sapi
26	30.4	13.9	162896	42	AC006441	AC006441 Homo sapien
27	30.2	13.9	107600	8	ATF16013	ATF16013
28	30	13.8	6150	12	DBPM1	DBPM1
29	29.8	13.7	4624	7	AC004770	AC004770
30	29.8	13.7	185035	11	AC004770	AC004770
31	29.8	13.7	5446	12	RATOSTP	RATOSTP
32	29.6	13.6	2537	4	AF026800	AF026800 Fundulus
33	29.6	13.6	2727	4	GSTRKA	GSTRKA
34	29.6	13.6	2738	4	GGU43396	GGU43396 Gallus galli
35	29.6	13.6	46791	11	AC004699	AC004699 Homo sapi
36	29.6	13.6	408	11	HSU88358	HSU88358 Homo sapien
37	29.6	13.6	182721	35	AC006507	AC006507 Mus muscu
38	29.4	13.5	777	5	A51866	A51866 Sequence 30
39	29.4	13.5	79612	11	AC004073	AC004073 Human Chr
40	29.4	13.5	128340	34	AC002419	AC002419 Homo sapi
41	29.2	13.4	192853	34	AC005096	AC005096 Homo sapi
42	29.2	13.4	124083	34	HSJ899816	HSJ899816 Homo sapi
43	29	13.3	3671	12	MMU88873	MMU88873 Mus musculu
44	29	13.3	156065	34	HS111B22	HS111B22
45	28.8	13.2	130690	1	D90901	D90901 Synecocyst

## ALIGNMENTS

RESULT 1	AF039400	3311 bp	RNA	PRI	15-DEC-1998
LOCUS	AF039400				
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA, complete cds.				
ACCESSION	AF039400				
NID	94009457				
VERSION	AF039400.1	GI:4009457			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.				
TITLE	Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins				
JOURNAL	Genomics 54 (2), 200-214 (1998)				
MEDLINE	99047526				
REFERENCE	2 (bases 1 to 3311)				
AUTHORS	Gruber A.D., Elble, R. and Pauli, B. U.				
TITLE	Direct Submission				

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
Location/Qualifiers

## FEATURES

source

1.3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1.3311  
/gene="hCLCA1"  
352..3096  
/gene="hCLCA1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="pid:g4009458"  
/db\_xref="gi:4009458"  
/translation="MGPKSSVFILILHLEGALSNSLIQLNNNGEIVVAIDPNP  
EDELIIQIDKNTQASLYLEATGKREYFNVAIIPEWTAKADYVRPLEYKNA  
DVLVAESTPGNDEPYTEMONGCEKGERIHLPDPLAKKLAEGPOGAPFHEMAH  
LRMGVDEYNDEKFIYLSNGRIQAVCSAGITGVNKKCOGSGCYTRKCFENKVTGL  
YKGECEVLSRQTEKASIMFAOHVDSIYECFTEQNHNRKAPNKNOKNCRSTWEVI  
RDSERFKTTPTQPPNPPTSLLQIGRIYCVLDKSGMATGRRLRLNQAQGLFL  
LQTVELSGMWGVTDSAAHVOSELIOINSQSDRDTLAKRLPAASGETSICGLSKA  
FTVIRKKTPTDSEIVLITDGEDNTISCFENVKOSGAIHTVALGPSAOLELELSK  
MTGLOIYASDOVONNGLIDAFGALSNGNAVORSIOLESGLTLONNGMNGVIY  
DSTVKDLFLITWTQPPQILLMPGSKOGFVDNKTMAKYLQIPGIAKMGVWY  
SLAASSOTLTLVTSRASNAITPTVTSKINKDTSKPSPLVYVANIROGASPIIRA  
SVTALISVNGKTVTLELDNGAGADTKDGVYSRFTYTDNGRSVAVRALGVN  
AARRRVIPQSGALYIPGWIENDEIQMNPPEIKNDVQKQVCFSTSSGSEFVAS  
DVNPAPIDLPFPGQITDLKAEIHGSLINLWTAFGDYDHGTAHRTIKRISTSLD  
LRDNFESLQVNTTALIPKANSSEVLEFKENITFENGTDLFLAIQAVDVLDKSEI  
SNIAVSLFIPQTPPETPPEDETSAPCPNHINSTITPGIHLIKIMKMWIGELQSLIA  
"

gene  
CDS

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 218; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 1.3e-64;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAACAATGGCCCTCATGATGCTTTGGGCGCTTTCATCAGAAATGAGCTGTCTCTCA 60  
DB 1743 GAACAATGGCCCTCATGATGCTTTGGGCGCTTTCATCAGAAATGAGCTGTCTCTCA 1802  
QY 61 GCGCTCCATCCAGCTTGAGAGTAAGGATTAACTCCAGAACAGCCAGTGTGATGATG 120  
DB 1803 GCGCTCCATCCAGCTTGAGAGTAAGGATTAACTCCAGAACAGCCAGTGTGATGATG 1862  
QY 121 CACAGTATGCTGAGACAGCAGCGTGGGAAAGAGACACTTGTCTTATCACCCTGAGAAC 180  
DB 1863 CACAGTATGCTGAGACAGCAGCGTGGGAAAGAGACACTTGTCTTATCACCCTGAGAAC 1922  
QY 181 GGAGCTCCCAATCTCTCTGCGATCCCGAGTGGAC 218  
DB 1923 GGAGCTCCCAATCTCTCTGCGATCCCGAGTGGAC 1960

RESULT 2  
LOCUS AF127036 2826 bp mRNA PRI 16-APR-1999  
DEFINITION Homo sapiens calcium-activated chloride channel protein 1 (CaCC1)  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
AUTHORS  
TITLE  
CHANNEL  
JOURNAL  
Unpublished  
2 (bases 1 to 2826)  
Agnel, M., and Culouscou, J.-M.  
Direct Submission  
Submitted (09-FEB-1999) Groupe Genomique, SYNHELABO, 10 Rue des  
Carrieres, Rueil-Malmaison 92500, France

FEATURES  
source  
1.2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1.2826  
/gene="CaCC1"  
5..2749  
/gene="CaCC1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD5487.1"  
/db\_xref="pid:g4585469"  
/db\_xref="gi:4585469"  
/translation="MGPKSSVFILILHLEGALSNSLIQLNNNGEIVVAIDPNP  
EDELIIQIDKNTQASLYLEATGKREYFNVAIIPEWTAKADYVRPLEYKNA  
DVLVAESTPGNDEPYTEMONGCEKGERIHLPDPLAKKLAEGPOGAPFHEMAH  
LRMGVDEYNDEKFIYLSNGRIQAVCSAGITGVNKKCOGSGCYTRKCFENKVTGL  
YKGECEVLSRQTEKASIMFAOHVDSIYECFTEQNHNRKAPNKNOKNCRSTWEVI  
RDSERFKTTPTQPPNPPTSLLQIGRIYCVLDKSGMATGRRLRLNQAQGLFL  
LQTVELSGMWGVTDSAAHVOSELIOINSQSDRDTLAKRLPAASGETSICGLSKA  
FTVIRKKTPTDSEIVLITDGEDNTISCFENVKOSGAIHTVALGPSAOLELELSK  
MTGLOIYASDOVONNGLIDAFGALSNGNAVORSIOLESGLTLONNGMNGVIY  
DSTVKDLFLITWTQPPQILLMPGSKOGFVDNKTMAKYLQIPGIAKMGVWY  
SLAASSOTLTLVTSRASNAITPTVTSKINKDTSKPSPLVYVANIROGASPIIRA  
SVTALISVNGKTVTLELDNGAGADTKDGVYSRFTYTDNGRSVAVRALGVN  
AARRRVIPQSGALYIPGWIENDEIQMNPPEIKNDVQKQVCFSTSSGSEFVAS  
DVNPAPIDLPFPGQITDLKAEIHGSLINLWTAFGDYDHGTAHRTIKRISTSLD  
LRDNFESLQVNTTALIPKANSSEVLEFKENITFENGTDLFLAIQAVDVLDKSEI  
SNIAVSLFIPQTPPETPPEDETSAPCPNHINSTITPGIHLIKIMKMWIGELQSLIA  
"

gene  
CDS

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 218; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 1.2e-64;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAACAATGGCCCTCATGATGCTTTGGGCGCTTTCATCAGAAATGAGCTGTCTCTCA 60  
DB 1396 GAACAATGGCCCTCATGATGCTTTGGGCGCTTTCATCAGAAATGAGCTGTCTCTCA 1455  
QY 61 GCGCTCCATCCAGCTTGAGAGTAAGGATTAACTCCAGAACAGCCAGTGTGATGATG 120  
DB 1456 GCGCTCCATCCAGCTTGAGAGTAAGGATTAACTCCAGAACAGCCAGTGTGATGATG 1515  
QY 121 CACAGTATGCTGAGACAGCAGCGTGGGAAAGAGACACTTGTCTTATCACCCTGAGAAC 180  
DB 1516 CACAGTATGCTGAGACAGCAGCGTGGGAAAGAGACACTTGTCTTATCACCCTGAGAAC 1575  
QY 181 GGAGCTCCCAATCTCTCTGCGATCCCGAGTGGAC 218  
DB 1576 GGAGCTCCCAATCTCTCTGCGATCCCGAGTGGAC 1613

RESULT 3  
LOCUS AB017156 2937 bp mRNA ROD 04-MAR-1999  
DEFINITION Mus musculus gob-5 mRNA, complete cds.  
ACCESSION AB017156



NID 93721911  
 VERSION AB017156.1 GI:3721911  
 KEYWORDS G0B-5.  
 SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (sites)  
 AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
 TITLE Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice  
 JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
 MEDLINE 9160866  
 REFERENCE 2 (bases 1 to 2937)  
 AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/Genbank databases. Tohru Komiyama, ERATO, JSR, Hirohashi Cell Configuration Project; 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan  
 (E-mail:tkom@ncp.jst.go.jp, Tel:81-298-47-7563, Fax:81-298-47-5226)  
 FEATURES  
 source  
 1. 2937  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_type="goblet cell"  
 /dev\_stage="adult"  
 /tissue\_type="intestine"  
 15..2756  
 /gene="gob-5"  
 15..2756  
 /gene="gob-5"  
 /codon\_start=1  
 /protein\_id="BAA33743.1"  
 /db\_xref="PID:01034712"  
 /db\_xref="PID:g3721912"  
 /db\_xref="GI:3721912"  
 /translation="MESLSPVFLILHLEGVLSLSLIQLNNNGEGIVAIIDHDP  
 EDEALIOHIDKMTQASPYLFGATGRFRFKNVAILIPESMRKPEYTKLETFKNA  
 DVLVSTSPLEDEPTEHIGAGEGKIRIHLTPDLAKLTQYGPORPTVEHVAH  
 FPMGVENENDEKFEYLSKRGQAVKCSAIIKGNVRGOGSCITNKCVIDRYTG  
 LYKNCVPPDPHONKASIMFNOMINVSVECTEKNHDEAPNONONCNARSTREY  
 IOESDFKOTTPMTAOPAPPTSLQIQRTICVLDKSGSLNDRLRKMAQSLF  
 LQIVQSGWGVMTFDSAAVQSEIKQNSADROLIKHLPTVSAGSTSCSGILRT  
 AFTVIRKKYPTDGESEIVLTDEDMTISSCFDLVKSGAIHITVALGPAAAELEDS  
 KFTGLOQYSSDOVONNGIVDAFPAALSSGNAIAQSIQLESVNLONNMNGSVI  
 VSGVGRDLFLITWTTHPTFTFIMPDSGEONGFLDTTRVAAYLOVPGTAKGVFK  
 VSGASQRTLTITVSRAATLPTPTTPVYNNKNTGKRPSPVYASIRGASPIIR  
 ASVTALIESVKNRTVLELNDGAGADATKNDGVSRFTAFDANGKISVKIMALGV  
 TSDRRAAPPKNRAMVIDGIMIDGVRNMPRPRTSYVDKOLCFSTRSSGSEVATN  
 VPAAPIDPLPFCQITDLKASIOGQNLVNLMTWAPGDYDHGRASNYIIRNSTSYD  
 LBDHNTSLQVNTGLIPKESASEEFELGNTGNGTDIFIAIOAVDKSNLKEI  
 SNIAVSVPINPDEPIPEDSPPCDDISINSTIPIGIVHLKIMKMLGMYVTGLGH"  
 BASE COUNT 860 a 718 c 693 g 666 t  
 ORIGIN  
 Query Match 71.7%; Score 156.2; DB 12; Length 2937;  
 Best Local Similarity 82.5%; Pred. No. 2.4e-43;  
 Matches 179; Conservative 0; Mismatches 38; Indels 0; Gaps 0;  
 Oy 1 GAACAATGGCCCTGATGATCTTTGGGCGCCCTTTCATCAGGAATGAGCTGTCTCA 60  
 Db 1409 GAACAATGGCTGTGATGATCTTTCGACGACATCTCTCAGGAATGCGCATCGCTCA 1468  
 Oy 61 GCGCTCATCTCAGCTGAGAGTAAGGATTAACCTCCAAAGACGACAGTGAATGG 120  
 Db 1469 GCACTCATCTCAGCTGAGAGTAAGGATTAATCTCAAAATAAGATGAATGG 1528  
 Oy 121 CACAGTATGCTGAGACAGCACCTGGGAAAGACACTTTTCTTATCCTGACCAAC 180  
 Db 1529 CTCAGTATGCTGAGACAGCTCGGTGGGCAAGACACTTGTTCCTTATCCTGACCAAC 1588

Oy 181 GCAGCTCCCAAAATCTTCTCGGATCCCACTGCA 217  
 Db 1589 GCATCTCTCACAATATTATCTGCGATCCCACTGCA 1625  
 RESULT 4  
 LOCUS AF039401  
 DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene, complete cds.  
 ACCESSION AF039401  
 NID 94009459  
 VERSION AF039401.1 GI:4009459  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 35278)  
 AUTHORS Gruber, A.D., Elble, R.C., Ji, H.L., Schreuer, K.D., Fuller, C.M. and Pauli, B.U.  
 TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
 JOURNAL Genomics 34 (2), 200-214 (1998)  
 MEDLINE 99047526  
 REFERENCE 2 (bases 1 to 35278)  
 AUTHORS Gruber, A.D., Elble, R. and Pauli, B.U.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
 FEATURES  
 source  
 1. 35278  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1p22-p31"  
 1..1617  
 /gene="hCLCA1"  
 1..33522  
 /gene="hCLCA1"  
 767..1265  
 /rpt\_family="L1 transposable element"  
 /rpt\_type="dispersed"  
 1587..1596  
 /gene="hCLCA1"  
 /note="putative"  
 1618..1708  
 /gene="hCLCA1"  
 /number=1  
 101n1.1618..1708,1962..2383,6667..6807,6979..7126,6966..9801,15454..15631,18591..18609,19783..20010,22253..24522..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)  
 /gene="hCLCA1"  
 /product="calcium-dependent chloride channel-1"  
 1962..2383  
 /gene="hCLCA1"  
 /number=2  
 101n1.2222..2383,6667..6807,6979..7126,6966..9801,15454..15631,18591..18609,19783..20010,22253..24522..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)  
 /gene="hCLCA1"  
 /codon\_start=1  
 /product="calcium-dependent chloride channel-1"  
 /protein\_id="AAC95429.1"  
 /db\_xref="PID:g4009460"  
 /translation="MGPKSSVFILHLLEGALSNSLIQLNNNGEGIVAIIDPNP  
 EDTELIOIKMTQASLYLFEATGRFRFKNVAILIPETWTKADYVRPKLEYKNA  
 DVLVAESTPRGNDPEYTPKMGKSGERHITPTPIAKGKIAEYRPGKAPVHNRH  
 LMGVFEINDEKFLSNGRIOAVKCSGILITINVKKCGGSCITRACITFNKVTGL"



QY	185	CCTCCCAATTCCTTCGGAGTCCCAATGGA	217
Db	1597	AAGCCAGAAATCTTCTCCAGCATCCAAAGGA	1629
RESULT	6		
LOCUS	AF001261		
DEFINITION	Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA,		
ACCESSION	AF001261		
NID	92623762		
KEYWORDS	AF001261.1 GI:2623762		
SOURCE	Bos taurus.		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.		
AUTHORS	1 (bases 1 to 3317) Eblie, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A., and Pauli, B.U.		
TITLE	Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel		
JOURNAL	Unpublished		
AUTHORS	2 (bases 1 to 3317) Eblie, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A., and Pauli, B.U.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA		
FEATURES	Location/Qualifiers		
source	1..3317 /organism="Bos taurus" /db_xref="taxon:9913" /cell_type="aortic endothelial cells" /clone="1" /tissue_type="Lung" 63..2780 /note="endothelial adhesion molecule; chloride channel" /codon_start=1 /product="Lu-ECAM-1" /protein_id="AAB86529.1" /db_xref="PID:92623763" /db_xref="GI:2623763" /translation="MYCLANVILFLTHLPKMKSSMVNLINNGYDVIAINSVPE DEKLEIKENKVEVTEASTYLPHFATRRRVYFRNVALILPMVWKSSEYFIPOESYDQAD VIVANPLKYGDDPYTLQYGRGCKGRXYIHTPNFLTNNEHIGSRVYFVHMAH, RMGFDEYNDQOPYISIRKNTIEATRSCTATGTGVNVEKCPGSGCTSLCRDSOQY LMEKCFELPKSOTAKESIMFMSLSVTEFCERHNEPAPMLQMKNGKSTWV IMNSVDQONTSPMTEENMPHPPTSLKSRVAVCYLDLDSGSSADRLPQONQAE LYLQVLEKGSLVGVTWFDVSAEIQNHILRTDNYVQITAKLPQVANGSTICREL KAGQVAILHSDQSTGSEIILLTDGENEINSCEDYKRSQALITHYALGDSAKLEL TLSMGTQGFREFFANKDITGLNARSIRSSRSGSITQOALQESKALITGRKRVNGTV PVDSVNDGFEFFVTWTLOKEEIVLODPKGRKYSPEDEKLNIRARLPQIATET GTWYSLINNASSOMLVTWYTTARSPTIPPVATAPAHMSQHTAHYSPMIVAVSOQ GFLPGLISVALIETEDGHQVTELEMDNGCRPTVNDGYSYFPTYDNGNGYSKIK VHAQARNNTARLNIRQONKLYIVPGYENKLIILNPPREPVKDLAKAKIEDSRILT SGSGFTVSGAPPEGNHPSVFPSPKITYDLEAFKEDYIQLSWTAGNVLDGKASVITI RISKSFVDRQEDFNATLVNTSNLIPREAGSENEFERPEHFRNENGTKKYISVQALIN EANLISVSHVQAKIFPLPEDSVHDLGTIKISEITLAILGLPMIFSEVF"		
BASE COUNT	1084 a 676 c 634 g 923 t		
ORIGIN			
Query Match	40.5%; Score 88.2; DB 3; Length 3317;		
Best Local Similarity	63.4%; Pred. No. 6.5e-20;		
Matches 135; Conservative	0; Mismatches 78; Indels 0; Gaps 0;		
QY	5	AATGCCCTCATTCATGCTTTGGGSCCTTTCATCAGGAATAGAGCTGCTCTCAGGCC	64
Db	1464	ACTGGCCCTTACTAAAGCTTTCAGTGAATTCATCTTAAGAGTGAAGCATCATCTAGAG	1523

Query Match 40.5%; Score 88.2; DB 3; Length 3288;







```
gene
361..2091
/map="76CD"
/gene="Deaf1"
/note="Deformed epidermal autoregulatory factor-1"
/allele=""
/db_xref="FlyBase:FBgn0013799"
361..2091
/gene="Deaf1"
/standard_name="Deformed epidermal autoregulatory factor-1"
/note="DNA-binding protein"
/codon_start=1
/product="dDeaf-1"
/protein_id="AAC47040.1"
/db_xref="PIR:G1209883"
/translation="MEQVDSSTELHLNRDLALADVYKEEYILESSHHHHHHH
LDTKVMATSSNDNSGSGAGGTGAGGNGGAVSVPLPIGSMITGTFVIT
TPDQLPPIKPMICVDNNGYLSGTYSMKNDKTYIQOQITPQGGGGANNAGTNT
TATNTGLNHDGSGNSHDSLTLEHAAGAGAGVGGGGTGGSSGSENSTON
EYFQIRKTYCALYRSKLSSGGRVYIKDWHPSEEHVGRGSSKDWRSIKY
GKSLQSLIDEGTLTPHATNCSCVCCDDEAGSAGPVPALPYRRRRNQDLSEN
SGPKRRNTSHSNNSNTNNNTSGGANNCYDTAAVAATAAYDENNFLSEEN
ITSKDEPMALNDLSTELVDSQMGNTYETEVVINDGSSIAVLDTSGSMKI
EHYCTMVKATNDFKRLNDMOSFERRIEVLQKEDAAVASMRVQVHADIDDPNIG
SLHNEITSAKKCANCRRLAECSLCRKTPICSEFCQKRDNNAHVECTRNPTTIO
QVWLIDDS"

BASE COUNT      666 a      664 c      630 g      454 t
ORIGIN

Query Match      15.0%; Score 32.6; DB 37; Length 2414;
Best Local Similarity 69.8%; Pred. No. 0.87;
Matches 44; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 94 CCTCCAGACAGCAGTGAATGATGCAGATGATGCAGACACCGTGGGAAGA 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1680 CCTGGACACAGCGCTGATGATGAAGACATGCAGACGCTGACTGACATGGCAAGC 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 154 CAC 156
    |||
DB 1740 CAC 1742

RESULT 14
AF041431      3197 bp      DNA      ROD      22-JAN-1998
LOCUS      Mus musculus Ls2 type intracisternal A-particle (IapIs2-2) gene,
DEFINITION      5'-flanking region, partial sequence.
ACCESSION      AF041431
NID      92801546
VERSION      AF041431.1 GI:2801546
KEYWORDS
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      Lueders, K.K., Mu, X. and Kuff, E.L.
TITLE      Changes in methylation of endogenous intracisternal A-particle
provided LTRs identify multiple genomic loci that are frequently
hypermethylated in transformed mouse cells
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 3197)
AUTHORS      Lueders, K.K.
TITLE      Direct Submission
JOURNAL      Submitted (06-JAN-1998) Laboratory of Biochemistry, NCI, NIH, 37
Convent Dr., Room Number 4D14, Bethesda, MD 20892-4225, USA
FEATURES
source
1..3197
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
```

```
gene
1..3197
/chromosome="6"
/map="near Kcnj2"
/clone="F.002F6"
/clone_lib="Prlinceton YAC library"
1..3197
/gene="IapIs2-2"
/note="Ls2 type intracisternal A-particle; 5'-flanking
region; sequences more proximal to IapIs2-2 are reported
in GenBank Accession Number AF019975"
1..93
/rpt_family="LINE"
/rpt_type="dispersed"
159..240
/rpt_family="LINE"
/rpt_type="dispersed"
315..473
/rpt_family="SINE/B2"
/rpt_type="dispersed"
2309..2364
/note="PABL A repeat"
2605..2679
/rpt_family="SINE/B4"
/rpt_type="dispersed"
2915..2978
/rpt_family="MER44C"
/rpt_type="tandem"

BASE COUNT      862 a      820 c      753 g      743 t      19 others
ORIGIN

Query Match      14.8%; Score 32.2; DB 12; Length 3197;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 76; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 23 TTGGGGCCCTTCATCAGGAATGAGCTGCTCAGCGGCTCCATCCAGCTTAGAG 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 TTTTGNCCCTCTTCAAGAGAGTGATGTTCTTCAGCGGCTGACACAGTATAGG 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 AAGGATTAACCTCCAGACAGCAGTGAATGACACACTGATCGTGAACAGCACC 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 134 CTGTTCTTCAGGCTACAAACATCCAGAGAGGGGTGCTGGAGAGTTTGAAGAGAA 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 GTGGAAAGACACTTGTCTTATCACT 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 194 AAGAGAGGGAATTAATTAATTAATTAAT 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AC004676      201175 bp      DNA      HTG      10-MAY-1998
LOCUS      Homo sapiens chromosome X clone BWXDI171, WORKING DRAFT SEQUENCE, 1
DEFINITION      ordered pieces.
ACCESSION      AC004676
NID      93126882
VERSION      AC004676.1 GI:3126882
KEYWORDS      HTG; HTGS_PHASE2.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS      Chen, E., Brownstein, B.H., States, D.J., Schlessinger, D. and
Mazzarella, R.
TITLE      Direct Submission
JOURNAL      Unpublished (1997)
REFERENCE      2 (bases 1 to 201175)
AUTHORS      Brownstein, B.H., States, D.J. and Mazzarella, R.
TITLE      Direct Submission
JOURNAL      Submitted (10-MAY-1998) Center for Genetics in Medicine, Box 8232,
Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA
COMMENT
Current status of this project is available at:
'http://www.idc.wustl.edu/cgm/seq_projects.html'
```





GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:11:37 ; Search time 3264.21 Seconds

(without alignments)  
246.497 Million cell updates/sec

Title: US-09-049-696-8

Perfect score: 253  
Sequence: 1 AACCAAGTGGTCATCATC.....GGCAGAGTATGTCGACAG 253

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBank: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pac: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_sc: \*  
14: gb\_sts: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_vl: \*  
18: em\_fun: \*  
19: em\_hcg: \*  
20: em\_hum1: \*  
21: em\_hum2: \*  
22: em\_in: \*  
23: em\_om: \*  
24: em\_or: \*  
25: em\_ov: \*  
26: em\_pat: \*  
27: em\_ph: \*  
28: em\_pl: \*  
29: em\_ro: \*  
30: em\_sts: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_vl: \*  
34: gb\_hcg1: \*  
35: gb\_hcg2: \*  
36: gb\_in1: \*  
37: gb\_in2: \*  
38: em\_ba1: \*  
39: em\_ba2: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query 8  
No. Score Match Length DB ID Description  
1 253 100.0 3311 11 AF039400 Homo sapi AF039400 Homo sapi

2	253	100.0	2826	42	AF127036	AF127036 Homo sapi
3	174.6	69.0	2937	12	AB017156	AB017156 Mus muscu
4	111.6	44.1	35278	11	AF039401	AF039401 Homo sapi
5	93.4	36.9	3604	9	AB026833	AB026833 Homo sapi
6	87	34.4	3317	3	AF001261	AF001261 Bos tauru
7	87	34.4	3388	3	AF001262	AF001262 Bos tauru
8	87	34.4	2820	3	AF001263	AF001263 Bos tauru
9	85.4	33.8	2884	3	BT036445	BT036445 Bos tauru
10	82	32.4	3022	12	AF047838	AF047838 Mus muscu
11	82	32.4	3471	12	AF052746	AF052746 Mus muscu
12	57.4	22.7	3415	42	AF043976	AF043976 Homo sapi
13	34	13.4	161657	11	AC005483	AC005483 Homo sapi
14	33.6	13.3	35412	7	SPAC478	SPAC478 Homo sapi
15	32.8	13.0	156736	42	AC006502	AC006502 Homo sapi
16	32.6	12.9	75777	9	HS276R20	HS276R20 Homo sapi
17	32.4	12.8	1484	1	MHAJ1656	MHAJ1656 Homo sapi
18	32.4	12.8	2354	1	MHAJ5054	MHAJ5054 Homo sapi
19	31.6	12.5	2489	4	XLGJUS15A	XLGJUS15A Homo sapi
20	31.6	12.5	3723	12	AF132478	AF132478 Homo sapi
21	31.6	12.5	5145	12	AF132481	AF132481 Homo sapi
22	31.4	12.4	197900	11	AC002531	AC002531 Homo sapi
23	31.2	12.3	8581	8	AF049110	AF049110 Homo sapi
24	31	12.3	124191	11	AC005153	AC005153 Homo sapi
25	30.8	12.2	1852	1	MHAJ5053	MHAJ5053 Homo sapi
26	30.8	12.2	1802	1	MHP50	MHP50 Homo sapi
27	30.8	12.2	262196	35	AC007513	AC007513 Homo sapi
28	30.6	12.1	81565	8	ATAC006072	ATAC006072 Arabidops
29	30.6	12.1	662	9	HUMRBP	HUMRBP Homo sapi
30	30.6	12.1	57991	11	AF013593	AF013593 Homo sapi
31	30.6	12.1	133153	34	HSDJ79963	HSDJ79963 Homo sapi
32	30.6	12.1	259037	35	AC005911	AC005911 Homo sapi
33	30.2	11.9	1803	1	MHAJ5055	MHAJ5055 Homo sapi
34	30.2	11.9	9890	2	AE001788	AE001788 Thermotog
35	30.2	11.9	128779	9	HS29X1	HS29X1 Homo sapi
36	30.2	11.9	146058	11	AC003668	AC003668 Homo sapi
37	30.2	11.9	187543	11	AC005678	AC005678 Homo sapi
38	30.2	11.9	151864	11	AF047825	AF047825 Homo sapi
39	30.2	11.9	159948	34	AC005103	AC005103 Homo sapi
40	30.2	11.9	158840	34	HS1186N24	HS1186N24 Homo sapi
41	30.2	11.9	31893	37	CELT22H9	CELT22H9 Homo sapi
42	30	11.9	4025	12	AF127798	AF127798 Rattus no
43	30	11.9	3812	12	AF132672	AF132672 Rattus no
44	30	11.9	194883	35	AC006430	AC006430 Homo sapi
45	29.8	11.8	12979	1	D90773	D90773 E.coli geno

## ALIGNMENTS

RESULT 1  
AF039400 3311 bp mRNA  
LOCUS AF039400  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hcccl1) mRNA,  
complete cds.  
ACCESSION AF039400  
NID 94009457  
VERSION AF039400.1 GI:4009457  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human CClA1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES Location/Qualifiers

Source

1.3311

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="1p22-p31"

/tissue\_type="small intestine"

gene

1.3311

/gene="hCLCA1"

352..3096

/note="transmembrane glycoprotein"

/codon\_start=1

/product="calcium-dependent chloride channel-1"

/protein\_id="AAC95428.1"

/db\_xref="pid:94009458"

/db\_xref="gi:4009458"

/translation="MGPKRSVFTIILHLBEGALNSLIQLNNNGYEGIVAVDPNP

EDETLIQIKDMVTOASLYLEATGKRFYKRWALLIPETWTKADYVRPLEYKNA

DVLVAESTPPGNDDEPTTEOMGNCGEKERHILTPDIFAGKLAEGPOGKAFFHEMAH

LRMGVDEYNDEKFTYLSNGRIQAVRCISAGITGVNVKCKOGSCYTRKCFNKVTGL

YKGECEVLOSROTEKASIMEAOHDSIYECTEONHNKEAPNOKNOKNLSRTEWEI

RDSERFKTTPMTPTOPPPNPTESLQIGORIVLYLDKSGSMATGRRLRNQAGOLF

LQTVELSGMGMVTFDSAAHVOSLTIQINSQSDROTILAKRLPAASGSGTSCGRLSA

FTVIRKRTPDGSEIVLITDGEDNTISGCFENVKSGAIHTVALGPSAOLELELSK

MTGGLQTVASDVQVONNGLIDAFGALSNGAVSORSIQLESKGLTLQNSOMNNTVLY

DSIVGKDTLELITWTTPOLILMDPSGOKGQGVVDNRKTMAYLIQIGIAKVTWKY

SVLQASQTLTLVTSRASNAITLPITVTSKTKRSKRPSPILVYVANTROGASPIIRA

SVALISVNGKTVTLLELDNGAGADATKDDGYSRFTYDNRKTSVYKRALGVN

AARRVITPOOSGALITPGIENDELOMNPPEIKRDYOHKOVCFSTSSGGSFVAS

DVNPALPDLPFGQITDLKAEIHGSGSLNLTWAPGDYDHGAHKYIIRISITILD

LKRDENSLQVNTTALIPKRNSEVEFLFKPENITFENGTDLFTAIQAVDKVLSSEI

SNIAVSLFIPPTPPETPSDETSAPCPNINHINSTITPGIHLIKIMMKWIGELQLSIA

BASE COUNT 1028 a 692 c 742 g 849 t

ORIGIN

Query Match 100.0%; Score 253; DB 11; Length 3311;

Best Local Similarity 100.0%; Pred. No. 1.6e-73;

Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACAAAGTGTGTCATCATCCACAGAGTGGCTTTGGGCCCCCTCTGCAGCTCAAGAACTAG 60

DB 1628 AACAAAGTGTGTCATCATCCACAGAGTGGCTTTGGGCCCCCTCTGCAGCTCAAGAACTAG 1687

QY 61 AGGAGCTGTCCAAATGACAGAGAGTTTACAGACATATGCTTCAGATTCAGAGTTGAGAAC 120

DB 1688 AGGAGCTGTCCAAATGACAGAGAGTTTACAGACATATGCTTCAGATTCAGAGTTGAGAAC 1747

QY 121 ATGGCCTCATGATGCTTTGGGCCCCCTTCATCAGAGAAATGAGAGCTGTCTCTCAGCGCT 180

DB 1748 ATGGCCTCATGATGCTTTGGGCCCCCTTCATCAGAGAAATGAGAGCTGTCTCTCAGCGCT 1807

QY 181 CCATCCAGCTTGAGAGTAAAGGATTAACCTCCAGAACCCAGAGTGAATGATGACAG 240

DB 1808 CCATCCAGCTTGAGAGTAAAGGATTAACCTCCAGAACCCAGAGTGAATGATGACAG 1867

QY 241 TGATCGTGACAG 253

DB 1868 TGATCGTGACAG 1880

RESULT 2

AF127036

LOCUS AF127036 2826 bp mRNA

DEFINITION Homo sapiens calcium-activated chloride channel protein 1 (caccl1)

ACCESSION AF127036

NID 94585468

VERSION AF127036.1 GI:4585468

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

gene

CDS

1.2826

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/tissue\_type="small intestine; colon"

/gene="caccl1"

5..2749

/note="bovine epithelial chloride channel homolog"

/codon\_start=1

/product="calcium-activated chloride channel protein 1"

/protein\_id="AAD25487.1"

/db\_xref="pid:94585469"

/db\_xref="gi:4585469"

/translation="MGPKRSVFTIILHLBEGALNSLIQLNNNGYEGIVAVDPNP

EDETLIQIKDMVTOASLYLEATGKRFYKRWALLIPETWTKADYVRPLEYKNA

DVLVAESTPPGNDDEPTTEOMGNCGEKERHILTPDIFAGKLAEGPOGKAFFHEMAH

LRMGVDEYNDEKFTYLSNGRIQAVRCISAGITGVNVKCKOGSCYTRKCFNKVTGL

YKGECEVLOSROTEKASIMEAOHDSIYECTEONHNKEAPNOKNOKNLSRTEWEI

RDSERFKTTPMTPTOPPPNPTESLQIGORIVLYLDKSGSMATGRRLRNQAGOLF

LQTVELSGMGMVTFDSAAHVOSLTIQINSQSDROTILAKRLPAASGSGTSCGRLSA

FTVIRKRTPDGSEIVLITDGEDNTISGCFENVKSGAIHTVALGPSAOLELELSK

MTGGLQTVASDVQVONNGLIDAFGALSNGAVSORSIQLESKGLTLQNSOMNNTVLY

DSIVGKDTLELITWTTPOLILMDPSGOKGQGVVDNRKTMAYLIQIGIAKVTWKY

SVLQASQTLTLVTSRASNAITLPITVTSKTKRSKRPSPILVYVANTROGASPIIRA

SVALISVNGKTVTLLELDNGAGADATKDDGYSRFTYDNRKTSVYKRALGVN

AARRVITPOOSGALITPGIENDELOMNPPEIKRDYOHKOVCFSTSSGGSFVAS

DVNPALPDLPFGQITDLKAEIHGSGSLNLTWAPGDYDHGAHKYIIRISITILD

LKRDENSLQVNTTALIPKRNSEVEFLFKPENITFENGTDLFTAIQAVDKVLSSEI

SNIAVSLFIPPTPPETPSDETSAPCPNINHINSTITPGIHLIKIMMKWIGELQLSIA

BASE COUNT 875 a 623 c 632 g 696 t

ORIGIN

Query Match 100.0%; Score 253; DB 42; Length 2826;

Best Local Similarity 100.0%; Pred. No. 1.5e-73;

Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACAAAGTGTGTCATCATCCACAGAGTGGCTTTGGGCCCCCTCTGCAGCTCAAGAACTAG 60

DB 1281 AACAAAGTGTGTCATCATCCACAGAGTGGCTTTGGGCCCCCTCTGCAGCTCAAGAACTAG 1340

QY 61 AGGAGCTGTCCAAATGACAGAGAGTTTACAGACATATGCTTCAGATTCAGAGTTGAGAAC 120

DB 1341 AGGAGCTGTCCAAATGACAGAGAGTTTACAGACATATGCTTCAGATTCAGAGTTGAGAAC 1400

QY 121 ATGGCCTCATGATGCTTTGGGCCCCCTTCATCAGAGAAATGAGAGCTGTCTCTCAGCGCT 180

DB 1401 ATGGCCTCATGATGCTTTGGGCCCCCTTCATCAGAGAAATGAGAGCTGTCTCTCAGCGCT 1460

QY 181 CCATCCAGCTTGAGAGTAAAGGATTAACCTCCAGAACCCAGAGTGAATGATGACAG 240

DB 1461 CCATCCAGCTTGAGAGTAAAGGATTAACCTCCAGAACCCAGAGTGAATGATGACAG 1520

QY 241 TGATCGTGACAG 253

DB 1868 TGATCGTGACAG 1880

DB 1521 TGATCGTGACAG 1533

RESULT 3  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS AB017156  
DEFINITION Mus musculus gob-5 mRNA, complete cds.  
ACCESSION AB017156  
NID 93721911  
VERSION AB017156.1 GI:3721911  
KEYWORDS GOB-5.  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (sites)  
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice  
JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
MEDLINE 99160866  
REFERENCE 2 (bases 1 to 2937)  
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru Komiyama, ERATO, JST, Hirohashi Cell Configuration Project; 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan  
(E-mail:tkom@hccp.jst.go.jp, Tel:81-298-47-7563, Fax:81-298-47-5226)

FEATURES  
source  
1. 2937  
location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15..2756  
/gene="gob-5"  
15..2756  
/codon\_start=1  
/protein\_id="BAA33743.1"  
/db\_xref="PID:d1034712"  
/db\_xref="PID:g93721912"  
/db\_xref="GI:3721912"  
/translation="MESLSPVFLILHLLEGLVSESLIQLNNNGYEGIVIAIDHVP  
EDEALIOHKDVTQASPILEATGKRFKRVAVLILPESSKAKPEYRPRKLEPKNA  
DYVSTTSPLENDPEYTEHICGCKGIRIHTTPFLAKKLTORGPODRVHEMAH  
FRMGVFNENDEKFTLSKGFQAVKCSAITGKQVRCQGGSCITNGKVIDRVG  
LYKDCVFPDPHONKASIMENONINSVEYTERKNOEAPNDQNCNRSYWEV  
IOESDFKQTPMTAPAPFSLQIGRIYCLVDKSGMLNDRRLRNMAQSLRF  
LLQVEQSGWMTVFDSAAVYOSLEKQDLSGADBLIKHLPTVSAGSTISCSGLRT  
AFTVIRKKRPTDGESEIVLLDGEDNTISCPFLVOSGAILHTVALGPAAKEELBOLS  
KMTGSLQYSSDQVONNGIYDAFALSSGNALIAHSDIOLERGVNLQNNQMNNSVI  
VDSVYKDTFLITITHTPTFTFMDPSVGEONGILDTTIVAILQYFGIKVGFPMK  
YSIQASSQTLITVTSRAASATLPITVTPVANKTKGFSPVTYASIRQASPILR  
ASVTALIESVNGKITVLELDNGAGADATKNDGVSRFTADANGRSYKIMALGV  
TSDROAAPKPRAMVIGDMIEDGRMNPPEPESYODKOLCFSSSGSFYATN  
VPAAPIDLPDPCOITDLKASIOGONLVNLTWATPGDDYDHGRASNTIRMTSTVID  
LADHNTSLQVNTTGLIPREASSEIFERELGNTFGNKTDFIALQAVDKSNLSEI  
SNIAVSVFIPAEPPIDPSTPPCDPDISINSTIGIHVLKIMRWLGMVOYTLGIH"

BASE COUNT 860 a 718 c 693 g 666 t

ORIGIN  
Query Match 69.0%; Score 174.6; DB 12; Length 2937;  
Best Local Similarity 80.6%; Pred.No.1.5e-47;  
Matches 204; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 1 AACAAAGTGTGCGCATCATCACACAGTGCCTTTGGGCGCTCGAGCGTCAAGAACTAG 60  
DB 1294 AGCAGAGCGGGGCGCATCATCATACAGTGGCCTGGGACCGCGCTCAAGAGCTTG 1353

OY 61 AGGAGCTGTCCAAAATGACAGGAGTTTACAGACATATGCTTCAGATCAATTCAGACACA 120  
DB 1354 AGCAGCTGTCCAAAATGACAGAGGCGCTGCAGACATCATCTCGATTCAGTTCCAGACACA 1413  
OY 121 ATGGCGCTCATATGATGCTTTGGGCGCTTTCATCATCGAATAAGGAGCTGCTCAGCGCT 180  
DB 1414 ATGGCTTTTATGCTTTTCGAGACGACACTCTCCTCAGGAATCCGGCGATCGCTCAGACT 1473  
OY 181 CCATCCAGCTTGAAGAGTAAGGATTAACTCCAGAACAGCCAGTGGATGAATGGCACAG 240  
DB 1474 CCATCCAGCTTGAAGAGTAAGGATTAACTCCAGAAATCAATGATGATGCTCAG 1533  
OY 241 TGATCGTGACAG 253  
DB 1534 TGATCGTGACAG 1546

RESULT 4  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS AF039401  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene, complete cds.  
ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber, A.D., Elble, R.C., Ji, H.L., Schreuer, K.D., Fuller, C.M. and Pauli, B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber, A.D., Elble, R. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

FEATURES  
source  
1. 35278  
location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
repeat\_region 767..1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="Calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,

24522..24628,26644..26859,27447..27708,28766..28936,  
31840..32079,32919..33310)  
/gene="hclca1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_18="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPKSSVFLLIHLLEGALSNSLIQLNNNGEIVVAIDPNV  
EDETLIQIKDMTOASLYLEATGKRRYFNKVALIETWTKADYRPLEYKNA  
DIVIAESTPPGNDPEYBOMGNGCGEGRHILTPETIAGKLAEGYGRKAFVEMH  
LRKGVFEDYNNDEKFTYLSNGRIQARCSAGITGTVYKCCGSCYTGRCPNVTGL  
YKGEFVLSRQTEKASIMFRQHYDSIVETCEONNKKKPNKONKCNIRSTWEVI  
RDSDEFKTTPMTQPPNPFTSLQIGRIYCLVDKSGSMATGRMLRMLQAGQLF  
LQTVELSGVMGVTFDSAAHYOSELIQNSGSDRDTLAKRLPAAAGSTSCGLRSA  
FTVIRKKYPTGSEIVLLDGEDNTISCFNEVKOSGAIHTVALGPAOAELELSK  
MTGLQYASDOVONNGLIDAFGALSNGNAVSOBSIOLESGITLIONSMNGTVIY  
DSYVKGDTFLITITPTOPQIILMDPSQKGGFVYDKNTKAYIQTIGIAKVGTVKY  
SLQASQTLITLVTSRASNATLPITVTSKINKDTSKPSPLVYANIRQASPIIRA  
SVTALIESVNGKIVYLELLDAGAGADAKDGVYSRYTTDINGRYSVKYRALGVA  
AARRVLIPOOGALYIPGMIENDEIOMNPPEIPKNDQVHKOVCFSTSSGSEFVA  
DVNPAPIDLPFGQITDLNAEIHGSSLINTLTPAGDYPDHGAHKYIIRISILD  
LRDNESLQVNTALIRKANSEVEFLFKENITFENGDTLTAIQADVDDKSEI  
SNIRVSLFIPQTPETPSPDETSAPCPNHINSTIGCIHLIKIMKMGIELQSTA

exon 6667..6807  
/gene="hclca1"  
/number=3  
6979..7126  
/gene="hclca1"  
/number=4  
9696..9801  
/gene="hclca1"  
/number=5  
15454..15631  
/gene="hclca1"  
/number=6  
18591..18809  
/gene="hclca1"  
/number=7  
19783..20010  
/gene="hclca1"  
/number=8  
22253..22427  
/gene="hclca1"  
/number=9  
24522..24628  
/gene="hclca1"  
/number=10  
26644..26859  
/gene="hclca1"  
/number=11  
27447..27708  
/gene="hclca1"  
/number=12  
28766..28936  
/gene="hclca1"  
/number=13  
31840..32079  
/gene="hclca1"  
/number=14  
32919..33522  
/gene="hclca1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN

Query Match 44.1%; Score 111.6; DB 11; Length 35278;  
Best Local Similarity 96.6%; Pred. No. 1,7e+26;  
Matches 114; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
CY 75 ATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTGACAGACAAAGGCGCTATTGAT 134

DB 24515 ATTTAGAGGAGTTTACAGACATATGCTTCAGATCAAGTTGACAGACAAAGGCGCTATTGAT 24574  
OY 135 GCTTTGGGGCCCTTTCATCAGAAATGAGCTGCTCTCAGCGCTCATCAGCTTG 192  
DB 24575 GCTTTGGGGCCCTTTCATCAGAAATGAGCTGCTCTCAGCGCTCATCAGCTTG 24632

RESULT 5  
AB026833  
LOCUS AB026833 3604 bp mRNA PRI 26-MAY-1999  
DEFINITION Homo sapiens mRNA for chloride channel protein, complete cds.  
ACCESSION AB026833  
NID 94887600  
VERSION AB026833.1 GI:4887600  
KEYWORDS chloride channel protein.  
SOURCE Homo sapiens adult corneal epithelium cDNA to mRNA,  
clone.lib:lambda ZAP clone:lambda 13802.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (sites)  
Itoh,R., Kawamoto,S., Kinoshita,S., Kawasaki,S. and Okubo,K.  
Isolation and characterization of chloride channel in human corneal  
epithelium  
Unpublished (1999)  
2 (bases 1 to 3604)  
Itoh,R., Kawamoto,S. and Okubo,K.  
Direct Submission  
Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases. Relko  
Itoh, Institute for Molecular and Cellular Biology, Osaka  
University, Yamada-oka 1-3, Suita, Osaka 565-0871, Japan  
(E-mail:reiko@imcb.osaka-u.ac.jp, Tel:81-6-6879-7992,  
Fax:81-6-6877-1922)

FEATURES  
source location/Qualifiers  
1..3604  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="lambda 13802"  
/clone.lib="lambda ZAP"  
/dex\_stage="adult"  
/map="1p32"  
/tissue\_type="corneal epithelium"  
42..2873  
/codon\_start=1  
/product="chloride channel protein"  
/protein\_id="BA077810.1"  
/db\_xref="PID:d1041578"  
/db\_xref="PID:g4887601"  
/db\_xref="GI:4887601"  
/translation="MTQRSIAGPICNLKFTVLVLSSELPFLGAGVQLDONGYGLL  
IAINPOVPEONLISNIKEMITASFLYFNATKRVFFRNKILIPATMRANNSKIK  
QSEYERKANVITDVGYGAGHGDPTLQVRCGGKGIHTPNFNLNLTAGSGRGR  
VFVHEAHNLWGVDEYINDKPYTINQONIKYTRCSDTIGTFCVCKGCGPOENCII  
SKLKEGCTITINSTONATASIMFOSLSSVEFCNSSTINQEPAPINQWCSIRSAM  
DVTITDSADFHSEPPMNGTELPPTFSLVQAGKYVCLVDVSSKMAEORLLQLODA  
AEFYLMQIVIEHTFVGIAFSDFSGKEIAQHLQINSNDRLVLSYLETFTYSAKTDISI  
CSGLKGEFVEVERKNGKAGYGMILVSGDGLNGCLPTVLSSGSIHISIALGSSAA  
PNLEELSRLTGGLKFTFPDISNSNSMIDASRGLSSGTDIFQOHIQLESGENVKPH  
QKATVTVNDTVGNDHFLVLTWQASGPEIILDDPDGRKTYTNNTFTNLTFRASLMI  
PGTAKPGHWITVTLNTHSHLQALKVYTSRASNAPPAIVEAVEVERDSLHFPHPVMI  
YANVKKQFYPILNATVATVETPEPGDVTLRLLDDGAGVAINDGIYSRFFSFAAN  
GRYSLKVHVNHSPISTPAHSIGSHMVPYGTANQINQMAPRKSGVGNDEERKAG  
FSRYSGGSGSVLGVPAGPDPVPPCKIIDLFAVYKEELTISMTAPGSDPQOGAT  
SVETRMKRSQNTQDDNNNAIIVNTSKRNPOQAGIRIETFFSPDISINSGEHPDNBT  
HESHRITVATRAMDRLNSQSAVSNIAQAPLFTPPNSDPVYARDYLILKGLVTLAMGLIG  
ITCLIVTHTHTLSRRKRKDRKENGTKLL"

BASE COUNT 1113 a 740 c 732 g 1017 t 2 others  
ORIGIN  
Query Match 36.9%; Score 93.4; DB 9; Length 3604;



/protein\_id="AAB86530.1"  
/db\_xref="PID:92623765"  
/db\_xref="GI:2623765"  
/translation="MYCLNVLFLTLHLPGMKSSMVLINNGYDGIYAIINPVE  
DEKLENIKEVYESTYLEFVATKRRVFRNYSILIPMTKSKSEYIIPROESYD  
VIVANPYLKGDDPYTLOYRGCEKGYIHFTPNLTLNPHYSGSRVPHMAHL  
RNGIDENVVDOPYISRKNTIARCSHITGINVFRKCGGSCITSLCRDSOTG  
LYEAKCTFLPKRSQTAKEIMFMSLSHSTETCTEHTNTEAPNLQNKMGKSTWY  
LINSVDFQNTSPMTENPPTHPTEFSLKSKQVCLVLDKSGMSAEDRLFGONQAE  
LYLQVIEKSLVGMVTEFDSVAEIONHLRITDDNVYOKITAKLPQVANGTSICRG  
KAGFOAIHSDDSTSGSEIILLDGEDNEINSCFEDVKSGAIHTIALGSAKELE  
TLSNMTGYRFFANKDITGLINAFSRISRSSTGOAIOLESKALKITGRNVGT  
PVSTVGNDFEVVWTIOKPEIVLODPKRYKTSDEKDKINISARLOIQTIAET  
GTPWYSLLNNHASSOMLYTVVTRARSPITPIVATRHMSOHTAHPSPMIVAYOVSQ  
GFLPVGLSVIAIETEDGHOVTLMDNGARDYKNDGYSRFTDYIGNRSLSK  
VHAQARNNTARLNROPQNVLYPGVYENGKIILNPREVKDILAKAKIEDFSRLT  
SGSFTVSGAPPGNHPVSFPPSKITDLEAKFREDYIOLSWTAPGNVLDKGKES"

BASE COUNT 1043 a 707 c 645 g 893 t

ORIGIN

Query Match 34.4%; Score 87; DB 3; Length 3288;  
Best Local Similarity 62.1%; Pred. No. 1.7e-18;  
Matches 157; Conservative 0; Mismatches 90; Indels 6; Gaps 1;

QY 1 AACAAAGTGGTCATCATCCACAGCTGGCTTTGGGCGCTCTGCAGCTCAAGAACTAG 60  
DB 1347 AACGAAGTGGTCATCATCCACAGCTGGCTTTGGGCGCTCTGCAGCTCAAGAACTAG 1406  
QY 61 AGAGAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATTCAGATTCAGACAA 120  
DB 1407 AGACATGTCAATATGACAGAGGATATC-----GTTTTTGGCAATTAAGACATAA 1460  
QY 121 ATGGCCTCATGATGCTTTGGGCGCTTCATCAGAGAAATGAGCTGTCTCAGCGCT 180  
DB 1461 CTGGCCTTACTAATGCTTTCAGTAGAATTTCACTAGAGTGAAGCATCCTCAGCAGG 1520  
QY 181 CCATCCAGCTTGAAGTAGAGGATTAACTCCAGAACGACGAGTGAATGAGCAGCAG 240  
DB 1521 CATTCAGTGTGAAGCAAGCCTTGAAATTAACAGAGAAAGAGTAAGGCGCAGCAG 1580  
QY 241 TGATCGTGACAG 253  
DB 1581 TGCTGTAGACAG 1593

RESULT 8  
AF001263 2820 bp mRNA MAM 19-NOV-1997  
LOCUS Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.  
ACCESSION AF001263  
NID 92623766  
VERSION AF001263.1 GI:2623766  
KEYWORDS Bos taurus.  
SOURCE Bos taurus  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 2820)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A., and Pauli, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
endothelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2820)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A., and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers

source 1. 2820  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone="3"  
/tissue="lung"  
/cell\_type="aortic endothelial cells"  
194. 2656  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86531.1"  
/db\_xref="PID:92623767"  
/db\_xref="GI:2623767"

CDS

BASE COUNT 940 a 565 c 560 g 755 t

ORIGIN

Query Match 34.4%; Score 87; DB 3; Length 2820;  
Best Local Similarity 62.1%; Pred. No. 1.7e-18;  
Matches 157; Conservative 0; Mismatches 90; Indels 6; Gaps 1;

QY 1 AACAAAGTGGTCATCATCCACAGCTGGCTTTGGGCGCTCTGCAGCTCAAGAACTAG 60  
DB 1482 AACGAAGTGGTCATCATCCACAGCTGGCTTTGGGCGCTCTGCAGCTCAAGAACTAG 1541  
QY 61 AGAGAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATTCAGATTCAGACAA 120  
DB 1542 AGACATGTCAATATGACAGAGGATATC-----GTTTTTGGCAATTAAGACATAA 1395  
QY 121 ATGGCCTCATGATGCTTTGGGCGCTTCATCAGAGAAATGAGCTGTCTCAGCGCT 180  
DB 1596 CTGGCCTTACTAATGCTTTCAGTAGAATTTCACTAGAGTGAAGCATCCTCAGCAGG 1655  
QY 181 CCATCCAGCTTGAAGTAGAGGATTAACTCCAGAACGACGAGTGAATGAGCAGCAG 240  
DB 1656 CATTCAGTGTGAAGCAAGCCTTGAAATTAACAGAGAAAGAGTAAGGCGCAGCAG 1715  
QY 241 TGATCGTGACAG 253  
DB 1716 TGCTGTAGACAG 1728

RESULT 9  
BT036445  
LOCUS BT036445 2984 bp mRNA MAM 09-FEB-1996  
DEFINITION Bos taurus calcium-activated chloride channel mRNA, complete cds.  
ACCESSION U36445  
NID 91184065  
VERSION U36445.1 GI:1184065  
KEYWORDS Bos taurus.  
SOURCE Bos taurus  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 2984)  
AUTHORS Cunningham, S.A., Awayda, M.S., Bublén, J.K., Ismailov, I.I.,  
Arrate, M.P., Berdiev, B.K., Benos, D.J., and Fuller, C.M.

TITLE Cloning of an epithelial chloride channel from bovine trachea  
JOURNAL J. Biol. Chem. 270 (32), 31016-31026 (1995)  
REFERENCE 96125078  
AUTHORS 2 (bases 1 to 2984)  
TITLE Benos, D.J.  
JOURNAL Direct Submission  
Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,  
University of Alabama at Birmingham, BHSB 706, Birmingham, AL  
35294, USA

FEATURES  
source Location/Qualifiers  
1. 2984  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19. 2730  
/note="epithelial chloride channel"  
/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="GI:1184066"  
/db\_xref="PIR:G1184066"

CDS  
translation="MPLRLVILFLTLHLPLGKSSMYLNNGYDGIYIAINPSVPE  
DEKLIONKEMVTEASTYLEFATKRRVYFRNLSILPMTWKSSEYIMPKROSYOAE  
VIVANPYLKHGDDPTTLOYGRGEGQYIHPTPNLITNLPYISGRAPVHEAHL  
RMGITDEYNGDPTTISRNTLEATRCSTHITGVIVACQGGSCITRCRVDSTGL  
YEAKCTIPEKQTARESIMFQSLASVETECTHVAEPNLONKNGSTMDVI  
NMSTDFONTSPTEMPNPPTQPFSLKSKORVCLDKSGMSSEDRFLRNQAL  
FLIIEKSLGVATVPDVAEIRNNLFTIDNYENTANLPOEANGSTICGLK  
AGFOALIOSQOSTSGSEIILFDGNEHSEIKOSGVIIHIALGPSAKRELT  
LSDMTGCHRFYANKINDITNANFSISSSGSIITQOTOLEKALAIIEKKVNTVP  
VDSITGNDTFEYVNTIKPELIDDPKGRKITSDFEDKLNISARLRIEGLTEG  
TWYSLLNNHASPILIVYTPRARSPTTPATAHMSQNAHPSPVIYAQVSQ  
FLPVGIVNTAIIETEDGHOVLELMDNAGADYKNGIYSRTDYRGNVYSLK  
HAERNTNARLSIROPKNALYIPGIENGKILNMPREVDLAKAEIDFSLTS  
GGSEFVSAPGNHPSVLPNKIIDLEAKFKEDHOLSWTANAVLDGKANSYIIRI  
SKSFIDKODPNALIVMTSSIKPKREASDENFERKPERPRENTNYIAQALNEA  
NITSFVSNIAQIKIIPMEDSVPLGRTISAINIATFLAMITLIV"

BASE COUNT 1027 a 583 c 561 g 813 t

ORIGIN

Query Match 33.8%; Score 85.4; DB 3; Length 2984;  
Best Local Similarity 61.7%; Pred. No. 5,8e-18;  
Matches 156; Conservative 0; Mismatches 91; Indels 6; Gaps 1;

QY 1 AACAAAGTGTGCATCATCACAGTGCCTTGGGCCCCCTGCAGCTCAAGAACTAG 60  
DB 1304 AACAAAGTGTGCATCATCACAGTGCCTTGGGCCCCCTGCAGCTCAAGAACTAG 1363  
QY 61 AGGAGCTGTCAAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTGACAGACA 120  
DB 1364 AGGAGCTGTCAAAATGACAGAGGTTTATGCGTTTATGCGTTCATTAAGACATATA 1417  
QY 121 ATGGCTCATATGATCTTTGGGGCCCTTCATCAGAGAAATGAGCTGTCTCACCCT 180  
DB 1418 ATGGCTCATATGATCTTTGGGGCCCTTCATCAGAGAAATGAGCTGTCTCACCCT 1477  
QY 181 CCATCCAGTTAGAGATGAGATTAACCTCCAGAAAGCCAGTGAATGAGCAGAC 240  
DB 1478 CTAATTCAGTTGGAAGCAAGCTTGCGCAATTCAGAAAAAGAAATGGTAAATGTTACGG 1537  
QY 241 TGATCGTGCAGAG 253  
DB 1538 TGCTGTGGATAG 1550

RESULT 10  
AF047838 3022 bp mRNA ROD 26-NOV-1998  
LOCUS Mus musculus calcium-sensitive chloride conductance protein-1  
DEFINITION (mMCAI) mRNA, complete cds.  
ACCESSION AF047838

NTD 93925280  
VERSION AF047838.1 GI:3925280  
KEYWORDS house mouse.  
SOURCE Mus musculus.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3022)  
AUTHORS Gandhi, R., Elble, R.C., Gruber, A.D., Schreier, K.D., Ji, H.T.,  
Fuller, C.M., and Pauli, B.U.  
TITLE Molecular and functional characterization of a calcium-sensitive  
chloride channel from mouse lung  
JOURNAL J. Biol. Chem. 273 (48), 32096-32101 (1998)  
REFERENCE 99041980  
MEDLINE 2 (bases 1 to 3022)  
AUTHORS Gandhi, R., Elble, R.C., Gruber, A.D., and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (12-FEB-1998) Vet. Pathology, Cornell Univ., WMC C4-146,  
Ithaca, NY 14856, USA

FEATURES  
source Location/Qualifiers  
1. 3022  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/tissue\_type="whole lung"  
1. 3022  
/gene="mMCAI"  
18. 2726  
/gene="mMCAI"  
/note="calcium-activated chloride channel: Alternative  
chloride channel to CFTR that may influence severity of  
cystic fibrosis in different tissues; mMCC-1"  
/codon\_start=1  
/product="calcium-sensitive chloride conductance  
protein-1"  
/protein\_id="AAC79982.1"  
/db\_xref="PIR:G3925281"  
/db\_xref="GI:3925281"

gene  
CDS  
translation="MPLQLVILFLTLHLNLTSSMYLNNSGYDGIYIAINPSVPE  
DERLIPSIKEMVTEASTYLEFASQGRVYFRNLSILPMTWKSSEYIMPKRESYDAD  
VIVADPHLHGDDPTTLOYGCGRGQYIHPTPNLITNLPYISGRAPVHEAHL  
RMGITDEYNGDPTTISRNTLEATRCSTHITGVIVACQGGSCITRCRVDSTGL  
YEAKCTIPEKQTARESIMFQSLASVETECTHVAEPNLONKNGSTMDVI  
KTSADPQNAFPMKGTAPPTPPTSLSKRRVYCLYDKSSMKREDRLTANQAEI  
VLTQIVKESVAGLVTEPDAHIONLYLKITSSDYQVITANLPQAGGTSICHQ  
AGPAITSSDSTSGSEIVLITDDEDMIRCFSPASVSGAIHITIALGPAARELET  
LSDMTGCHRFYANKINDITNANFSISSSGSIITQOTOLEKALAIIEKKVNTVP  
VDSITGNDTFEYVNTIKPELIDDPKGRKITSDFEDKLNISARLRIEGLTEG  
TWYSLLNNHASPILIVYTPRARSPTTPATAHMSQNAHPSPVIYAQVSQ  
FLPVGIVNTAIIETEDGHOVLELMDNAGADYKNGIYSRTDYRGNVYSLK  
HAERNTNARLSIROPKNALYIPGIENGKILNMPREVDLAKAEIDFSLTS  
GGSEFVSAPGNHPSVLPNKIIDLEAKFKEDHOLSWTANAVLDGKANSYIIRI  
SKSFIDKODPNALIVMTSSIKPKREASDENFERKPERPRENTNYIAQALNEA  
NITSFVSNIAQIKIIPMEDSVPLGRTISAINIATFLAMITLIV"

BASE COUNT 923 a 710 c 661 g 728 t

ORIGIN

Query Match 32.4%; Score 82; DB 12; Length 3022;  
Best Local Similarity 61.3%; Pred. No. 7,8e-17;  
Matches 152; Conservative 0; Mismatches 90; Indels 6; Gaps 1;

QY 6 AGTGTGCCATCATCACAGTGCCTTGGGCCCCCTGCAGCTCAAGAACTAGAGAG 65  
DB 1308 AGTGTGCCATCATCACAGTGCCTTGGGCCCCCTGCAGCTCAAGAACTAGAGAG 1367  
QY 66 CTGTCCAAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAAATGAGC 125  
DB 1368 CTGTCCAAAATGACAGAGGTTTCTTCTATGCAACAAAGACT-----AACAGC 1421  
QY 126 CTAATTCAGTTTGGGGCCCTTCATCAGGAAATGAGAGCTGTCTCAGGCTGCATC 185  
DB 1422 CTAATTCAGTTTGGGGCCCTTCATCAGGAAATGAGAGCTGTCTCAGGCTGCATC 1481

QY 186 CAGCTGAGAGTAAGGATTAAACCTCCAGAACGCCAGTGAATGGACAGTATC 245  
|||||  
Db 1482 CAGTTGGAGAGCAACCTTCGATGTCAGAGCAGGCGCATGATTAACGGTACAGTACCT 1541  
QY 246 GTGGACAG 253  
|||||  
Db 1542 CTGGACAG 1549

RESULT 11  
AF052746 3471 bp mRNA ROD 08-APR-1999  
LOCUS Mus musculus chloride channel cacc mRNA, complete cds.  
DEFINITION AF052746  
ACCESSION AF052746  
NID 93560546  
VERSION AF052746.1 GI:3560546  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS Romio, L., Musante, L., Cinli, R., Serl, M., Moran, O., Zegarra-Moran, O.  
and Galletta, L.J.V.  
TITLE Characterization of a murine gene homologous to the bovine Cacc  
chloride channel  
JOURNAL Gene 228 (1-2), 181-188 (1999)  
MEDLINE 99173882  
REFERENCE 2 (bases 1 to 3471)  
AUTHORS Galletta, L.J.V.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto  
Giannina Gaslini, Genova I-16148, Italy  
FEATURES  
source  
1. 3471  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/db\_xref="dbEST:AA245677"  
47..2752  
/codon\_start=1  
/product="chloride channel cacc"  
/protein\_id="AAC35003.1"  
/db\_xref="PID:93560547"  
/db\_xref="GI:3560547"  
/translation="MVRSLQVLLPLTLHLONTSSMWHLSNGEYVIAINISPE  
DERLPSIKEMVTOASTYLEFASGGRVFNISILVPTWRSKPEYLPKRESYDAD  
VIVADPHLOHGDDEPYTLOYGCGRGQYIHPTPNLITDNLRIYGRPRVHBMAL  
RMGVDEYNQPYMSRKNTIATRCSTRITGVNVNCHERGNCVTACRDRSTRL  
YEPKCTEIPDKIOTAGASIMEMQNLNSVECTENHNNAEAPNLONKCNRSWDYI  
KTSADFONAPPMRGTEAPPTPEFSLKSRRLVCLVLDKSSMCKEDRLIRMONAAL  
YLIOIVKESWGLVTPDPSAHIONYLKITSSTSSDOKITANLEQASGGSTHGLQ  
AGPRAITSSDOSTSGSEVLTLDDEDNIRSCFVSRSGAIHTIALGPAAELELT  
LSDTGGRLFRFANDLSLDAFSTRISSTGVSQAOLQESKAFDVRAGMNGITVP  
LDSTVNDTEFEVITWVKKEPIIQDPRGKRYTSDODKLNRSARLQIPGAEFG  
PTWISITGKSQILITMTVTRASPTMEPLATAMSGSTAYSPSMIYARVQOGL  
PVGANVTALIEAHGHQVTELEMDNGADTAVNDGITYTFYFDGNSNSYISKVY  
OAOBKTRLSLRKNSKSLIDGYENKRYINPRPDYQEALEATYEDRRRTYSGGS  
FYSGAPDGDHARVFPSPSKYTDLEAFETGDIHLTATPAKVLNDRARHYITIRMSQ  
HPDLQEDFNATLVNASSLIPKESKRETFKPEFPIKANGIQLVIAIQADNEASL  
TSEVSNTIAQAKLTSLSDSISALGDISICMTWIGLTVIFNSILN"

BASE COUNT 1096 a 777 c 736 g 862 t  
ORIGIN

Query Match 32.4%; Score 82; DB 12; Length 3471;  
Best Local Similarity 61.3%; Pred. No. 8e-17;  
Matches 152; Conservative 0; Mismatches 90; Indels 6; Gaps 1;

QY 6 AGTGTCATCACCACACAGTGCCTTTGGGCGCTTCGACGCTCAAGAACTAGAGAG 65  
Db 1337 AGCGGTGCATCATCCACACCATCTGCTGGGCGCTTCGCGCGAAGCTGAGACT 1396

QY 66 CTGTCCAAAATGACAGAGAGGTTTACAGACATATGCTTCAGATCAAGTTAGAACATGCC 125  
|||||  
Db 1397 CTGTGGACATGACAGAGGCGCTTCGTTTCATGCAACAAAGACT-----AAACAGC 1450

QY 126 CTCATGATGCTTTGGGCGCTTCATCAGAAATGAGAGCTGTCTCAGCGCTCCATC 185  
|||||  
Db 1451 CTATGAGATGCTTCAGATGATTCATCAAGTGGCAGCGCTCCACAGCGCTCG 1510

QY 186 CAGCTGAGAGTAAGGATTAAACCTCCAGAACGCCAGTGAATGGACAGTATC 245  
|||||  
Db 1511 CAGTTGGAGAGCAACCTTCGATGTCAGAGCAGGCGCATGATTAACGGTACAGTACCT 1570

QY 246 GTGGACAG 253  
|||||  
Db 1571 CTGGACAG 1578

RESULT 12  
AF043976 3415 bp mRNA PRI 07-APR-1999  
LOCUS Homo sapiens Clca homolog (hclca3) mRNA, complete cds.  
DEFINITION AF043976  
ACCESSION AF043976  
NID 94572288  
VERSION AF043976.1 GI:4572288  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Gruber, A.D. and Pauli, B.U.  
TITLE Molecular cloning and biochemical characterization of a truncated,  
secreted member of the human family of Ca2+-activated Cl- channels  
JOURNAL Biochim. Biophys. Acta 1444 (3), 418-423 (1999)  
MEDLINE 99196715  
REFERENCE 2 (bases 1 to 3415)  
AUTHORS Gruber, A.D., Elble, R.C. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1. 3415  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..3415  
/gene="hclca3"  
19..807  
/gene="hclca3"  
/product="Clca homolog"  
/codon\_start=1  
/protein\_id="AA23734.1"  
/db\_xref="PID:94572289"  
/db\_xref="GI:4572289"  
/translation="MVFSLKSLVLLSPVLSKSLVLTNNNGYDGIYAIINISPE  
DEKIONIKEMVTEASTLEFATKQAVRNYSILPMYKSKSEVLLIKROYDAD  
VIVADLYKGDDEPYTLOYGCGGDRQYIHPTPNLITDNLRIYGRPRVHBMAL  
RMGVDEYNQPYMSRKNTIATRCSTRITGVNVNCHERGNCVTACRDRSTRL  
YEPKCTEIPDKIOTAGASIMEMQNLNSVECTENHNNAEAPNLONKCNRSWDYI  
KTSADFONAPPMRGTEAPPTPEFSLKSRRLVCLVLDKSSMCKEDRLIRMONAAL  
YLIOIVKESWGLVTPDPSAHIONYLKITSSTSSDOKITANLEQASGGSTHGLQ  
AGPRAITSSDOSTSGSEVLTLDDEDNIRSCFVSRSGAIHTIALGPAAELELT  
LSDTGGRLFRFANDLSLDAFSTRISSTGVSQAOLQESKAFDVRAGMNGITVP  
LDSTVNDTEFEVITWVKKEPIIQDPRGKRYTSDODKLNRSARLQIPGAEFG  
PTWISITGKSQILITMTVTRASPTMEPLATAMSGSTAYSPSMIYARVQOGL  
PVGANVTALIEAHGHQVTELEMDNGADTAVNDGITYTFYFDGNSNSYISKVY  
OAOBKTRLSLRKNSKSLIDGYENKRYINPRPDYQEALEATYEDRRRTYSGGS  
FYSGAPDGDHARVFPSPSKYTDLEAFETGDIHLTATPAKVLNDRARHYITIRMSQ  
HPDLQEDFNATLVNASSLIPKESKRETFKPEFPIKANGIQLVIAIQADNEASL  
TSEVSNTIAQAKLTSLSDSISALGDISICMTWIGLTVIFNSILN"

BASE COUNT 1149 a 625 c 630 g 1011 t  
ORIGIN

Query Match 22.7%; Score 57.4; DB 42; Length 3415;  
Best Local Similarity 63.3%; Pred. No. 1.e-08;  
Matches 88; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 115 AGACATGCGCCATGATGCTTTGGGCGCTTCATCAGAAATGAGAGCTGTCTC 174  
Db 1469 ACATTAATGCGCTTATGATGCTTTTCAGCAAAATTCATCAGAGTGGACATCTCTC 1528

QY 175 AGCGTCCATCAGAGTGAAGGATTAAACCTCCAGAACGCCAGTGAATG 234  
Db 1529 AGCAGCTCTTCAGTTGGAAGATAAATCTTGATATCCAGCGAAGAAATGATAATG 1588



QY 235 GCACAGTGCCTGGACAG 253  
 1 ||||| ||||| ||  
 Db 1589 GTACAGTGCCTGGATG 1607

RESULT 13  
 AC005483/C 161667 bp DNA PRI 07-DEC-1998  
 LOCUS Homo sapiens PAC clone D0806A17 from 7p13-p14, complete sequence.  
 AC005483  
 AC005483  
 NID 93980526  
 VERSION AC005483.1 GI:3980526  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 161667)  
 Hou.S., McQuerry, Y., Yeakum, M. and Osborn, A.  
 TITLE The sequence of Homo sapiens PAC clone D0806A17  
 JOURNAL Unpublished (1998)  
 REFERENCE 2 (bases 1 to 161667)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-AUG-1998) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 3 (bases 1 to 161667)  
 Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-DEC-1998) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Dec 8, 1998 this sequence version replaced gi:3419854.  
 SUBMITTED BY: WUGSC  
 Genome Sequencing Center  
 Department of Genetics  
 Washington University  
 St. Louis MO 63108, USA  
 http://genome.wustl.edu/gsc  
 mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see  
 http://www.nhgri.nih.gov/DIR/GRB/CHR7, send  
 mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.  
 The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.

VECTOR: pCYPAC2  
 NEIGHBORING SEQUENCE INFORMATION:  
 The actual start of this clone is at base position 1 of D0806A17,  
 actual end is at 161667 of D0806A17.  
 Location/Qualifiers

FEATURES	source
1. 161667	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="7"	
/map="7p13-p14"	
/clone_lib="RPCI-4"	
/clone="D0806A17"	
176..495	
/rpt_family="L2"	
520..592	
/rpt_family="L2"	
806..999	
/rpt_family="L2"	
1587..1697	
/rpt_family="MERL_type"	
2042..2111	
/rpt_family="MIR"	
3790..3834	
/rpt_family="POLY_A"	
3840..4151	
/rpt_family="Alu"	
5143..5165	
/rpt_family="POLY_A"	
5573..5758	
/rpt_family="L1"	
5928..5967	
/rpt_family="AT-rich"	
5977..6103	
/rpt_family="MIR"	
6329..8269	
/rpt_family="L1"	
8270..8359	
/rpt_family="Mariner"	
8360..8764	
/rpt_family="L1"	
9144..9197	
/rpt_family="L2"	
9757..9851	
/rpt_family="(TTAA)n"	
9862..11819	
/rpt_family="L1"	
11820..12134	
/rpt_family="Alu"	
12135..12581	
/rpt_family="L1"	
12582..12620	
/rpt_family="AT-rich"	
12697..13033	
/rpt_family="MER4-group"	
13467..13691	
/rpt_family="MIR"	
13713..13874	
/rpt_family="MER53"	
13898..13923	
/rpt_family="(TGAA)n"	
13926..14304	
/rpt_family="L2"	
14534..14598	
/rpt_family="MERL_type?"	
14626..14691	
/rpt_family="L2"	
15484..15619	
/rpt_family="MERL_type"	
17051..17077	
/rpt_family="AT-rich"	
17866..18169	
/rpt_family="Alu"	
19068..19548	
repeat_region	







repeat\_region complement(21591..22303)  
 /rpt\_family="LIMEc" .22303)  
 repeat\_region complement(22306..22668)  
 /rpt\_family="THE1B"  
 repeat\_region 22735..22986  
 /rpt\_family="Alusg1"  
 repeat\_region complement(23189..23621)  
 /rpt\_family="LIMEc" .23621)  
 repeat\_region complement(23999..24068)  
 /rpt\_family="MLT1I"  
 repeat\_region complement(24448..24927)  
 /rpt\_family="LTR3"  
 repeat\_region complement(25013..25500)  
 /rpt\_family="L2"  
 repeat\_region complement(25532..25621)  
 /rpt\_family="L2"  
 repeat\_region 26921..26942  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(27281..27498)  
 /rpt\_family="Alusg/x"  
 repeat\_region complement(27499..27816)  
 /rpt\_family="AluYD8"  
 repeat\_region 28975..29043  
 /rpt\_family="AT\_rich"  
 repeat\_region 29157..29252  
 /rpt\_family="AT\_rich"  
 repeat\_region 29755..29833  
 /rpt\_family="AT\_rich"  
 repeat\_region 30114..30136  
 /rpt\_family="MLT1I"  
 repeat\_region 30114..30136  
 /rpt\_family="CATARA)n"  
 repeat\_region 31537..31665  
 /rpt\_family="(TA)n"  
 repeat\_region 31791..31934  
 /rpt\_family="(TATARA)n"  
 repeat\_region 32731..32759  
 /rpt\_family="AT\_rich"  
 repeat\_region 33207..33235  
 /rpt\_family="AT\_rich"  
 repeat\_region 33671..33723  
 /rpt\_family="AT\_rich"  
 repeat\_region 34354..34374  
 /rpt\_family="AT\_rich"  
 repeat\_region 34488..34508  
 /rpt\_family="AT\_rich"  
 repeat\_region 35006..35034  
 /rpt\_family="AT\_rich"  
 repeat\_region 37107..37574  
 /rpt\_family="LTR53"  
 repeat\_region 38023..38043  
 /rpt\_family="AT\_rich"  
 repeat\_region 39382..39477  
 /rpt\_family="MER5A"  
 repeat\_region 41196..41579  
 /rpt\_family="MST7A"  
 repeat\_region 42420..43326  
 /rpt\_family="L1PA2"  
 repeat\_region 43573..45117  
 /rpt\_family="L1MCB"  
 repeat\_region complement(45294..45813)  
 /rpt\_family="LTR38"  
 repeat\_region 45828..46421  
 /rpt\_family="L1M4"  
 repeat\_region complement(46431..46761)  
 /rpt\_family="MER39B"

Query Match 13.0%: Score 32.8; DB 42; Length 156736;  
 Best Local Similarity 50.6%: Pred. No. 2.9; Mismatches 77; Indels 0; Gaps 0;

Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
 QY 86 TTTACAGACATATGCTTCAGATCACTGATGAGAAATGAGGCTTCATGATGCTTTGGGGC 145  
 DB 115754 TTTAATATCAACCTTCATAGGCGGTTCCTTCAAGAGCGCGTGTGAGCTCCCTTCAGC 115813

QY 146 CCTTCATCAGAAATGAGACTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATT 205  
 DB 115814 ACTTCTGATGTAAGAAAGCCTTCCCAACAATGCTACACTTTCAAAAGCAAGGAAT 115873  
 QY 206 AACCTTCAGAACAGCCAGTGTGATGATGACAGT 241  
 DB 115874 AACCTTCGATCATCTCTCTCCATATAGCCCGT 115909

Search completed: August 5, 1999, 23:12:11  
 Job time: 7614 sec

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:40 ; Search time 3059.95 Seconds

(without alignments)  
163.091 Million cell updates/sec

Title: US-09-049-696-8

Sequence: 1 AACAAAGTGTGCATCANC.....GGCACAGTATCTGTGCAGAG 253

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	7.1	452	23	H52536	H52536 yq81f08.s1
2	18	7.1	382	30	AA243037	AA243037 zt24g05.s
3	18	7.1	434	38	AA771269	AA771269 vt17d09.r
4	17	6.7	314	20	M78894	M78894 EST01042.Su
5	17	6.7	272	20	T16657	T16657 NIB1651.Nor
6	17	6.7	263	20	T32493	T32493 EST49387.Hu
7	17	6.7	324	20	T35391	T35391 EST83977.Hu
8	17	6.7	313	20	T52594	T52594 yb21c03.r1
9	17	6.7	318	21	R09245	R09245 yf26g03.s1
10	17	6.7	311	23	D52818	D52818 HUM089B02B
11	17	6.7	333	23	L38062	L38062 BNAF0409E.M
12	17	6.7	365	24	N41057	N41057 yy53g06.s1
13	17	6.7	313	25	N66420	N66420 yz40c05.s1
14	17	6.7	474	25	N93765	N93765 zb64c07.s1
15	17	6.7	377	26	W44889	W44889 zc05g08.r1
16	17	6.7	410	26	W86876	W86876 zh63g08.s1
17	17	6.7	420	26	W86892	W86892 zh60d08.s1
18	17	6.7	413	27	AA046043	AA046043 zk76d08.s
19	17	6.7	365	27	W91192	W91192 mf72d07.r1
20	17	6.7	437	28	AA064183	AA064183 mj64f03.r
21	17	6.7	454	28	AA114065	AA114065 zn65b08.r
22	17	6.7	384	29	AA125957	AA125957 zn27e08.r
23	17	6.7	411	29	AA126011	AA126011 z185f02.s
24	17	6.7	442	29	AA132749	AA132749 zo21a06.s
25	17	6.7	213	29	AA137598	AA137598 mq86f12.r
26	17	6.7	565	29	AA167671	AA167671 zq39c07.r
27	17	6.7	191	30	AA269100	AA269100 K14SLB198
28	17	6.7	258	31	AA320402	AA320402 EST272752
29	17	6.7	242	32	AA348184	AA348184 EST35778
30	17	6.7	490	34	AA479093	AA479093 zv17e02.s
31	17	6.7	320	34	AA483422	AA483422 ne39h01.s
32	17	6.7	417	34	AA533289	AA533289 nj66f08.s
33	17	6.7	545	35	AA538001	AA538001 v13fg07.r
34	17	6.7	621	35	AA572759	AA572759 n119a01.s
35	17	6.7	268	35	AA579524	AA579524 n137g12.s
36	17	6.7	340	35	AA580059	AA580059 n179f07.s
37	17	6.7	321	35	AA588372	AA588372 nf40g01.s
38	17	6.7	324	36	AA622058	AA622058 ng54a02.s
39	17	6.7	431	36	AA622446	AA622446 ng40c08.s
40	17	6.7	427	36	AA642566	AA642566 ng73c02.s
41	17	6.7	558	37	AA695084	AA695084 gm02113.s
42	17	6.7	465	37	AA696591	AA696591 gm07964.5
43	17	6.7	553	37	AA697266	AA697266 h101084.5
44	17	6.7	463	37	AA705345	AA705345 z189c12.s
45	17	6.7	415	50	AT684786	AT684786 wa73f08.x

#### ALIGNMENTS

RESULT 1  
LOCUS H52536 452 bp mRNA  
DEFINITION yq81f08.s1 Soares fetal liver spleen INF1s Homo sapiens cDNA clone  
ACCESSION H52536  
IMAGE:202315 3', mRNA sequence.  
NID 9992377  
VERSION H52536.1 GI:992377

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 452)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
Insert Size: 937  
High quality sequence stops: 338  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 937 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 338.  
Location/Qualifiers  
1. 452  
/organism="Homo sapiens"  
/db\_xref="GDB:3771346"  
/db\_xref="taxon:9606"  
/clone="IMAGE:202215"  
/clone\_1lb="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGAGAGATTAAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 117 a 95 c 87 g 147 t 6 others  
ORIGIN

Query Match 7.1%; Score 18; DB 23; Length 452;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AGCTGTCCTCAAAATGACAG 81  
|||||  
DB 62 AGCTGTCCTCAAAATGACAG 79

RESULT 2  
AA243037 382 bp mRNA EST 11-MAR-1998  
LOCUS 2124905.s1 Striatogene NT2 neuronal precursor 937230 Homo sapiens  
DEFINITION cDNA clone IMAGE:664376 3' similar to gb:120868 60S RIBOSOMAL  
PROTEIN L4 (HUMAN);, mRNA sequence.  
ACCESSION AA243037  
NID 91873833  
VERSION AA243037.1 GI:1873833  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 382)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krieman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R., and Wilson, R.  
TITLE WashU-Merck human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On May 5, 1995 this sequence version replaced gi:798381.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1485 Std Error: 0.00  
Seq primer: -41ml3 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 382  
/organism="Homo sapiens"  
/db\_xref="GDB:5426364"  
/db\_xref="taxon:9606"  
/map="21"  
/clone="IMAGE:664376"  
/clone\_1lb="Striatogene NT2 neuronal precursor 937230"  
/dev\_stage="Ntera-2 neuroepithelial cells"  
/issue\_type="neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: brain; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Uninduced, exponentially growing neuroepithelial cells (Ntera-2/cl.DI). Average insert size: 1.0 kb; 3' -3' adaptor sequence: 5' GATTGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT 72 a 102 c 90 g 118 t  
ORIGIN

Query Match 7.1%; Score 18; DB 30; Length 382;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 AGGAGCTGTCCTCAAAATGA 78  
|||||  
DB 35 AGGAGCTGTCCTCAAAATGA 52

RESULT 3  
AA771269 434 bp mRNA EST 29-JAN-1998  
LOCUS v17d09.r1 Barstead mouse myotubes MFLRB5 Mus musculus cDNA clone  
DEFINITION IMAGE:1163345 5', mRNA sequence.  
ACCESSION AA771269  
NID 92823080  
VERSION AA771269.1 GI:2823080  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
REFERENCE 1 (bases 1 to 444)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R., and Waterston, R.  
TITLE The WashU-HMT Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407528.





BASE COUNT 58 a 64 c 55 g 95 t  
with spinal muscular atrophy. The library was normalized as described elsewhere.  
ORIGIN

Query Match 6.7% Score 17; DB 20; Length 272;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 AGAAGTAGAGAGCTGT 69  
|||||  
Db 191 AGAAGTAGAGAGCTGT 207

RESULT 6  
T32493 263 bp mRNA EST 06-SEP-1995  
LOCUS EST49387 Human Gall bladder Homo sapiens CDNA 3' end similar to  
DEFINITION None, mRNA sequence.  
ACCESSION T32493  
NID 9614591  
VERSION T32493.1 GI:614591  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 263)  
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,  
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S.,  
Kelley,J.M., Klinek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,  
Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,  
Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,  
Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R.,  
Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,  
Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrite,A.,  
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,  
Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,  
Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wang,J., Xu,C.,  
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,  
Hasselaine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.

Initial Assessment of Human Gene Diversity and Expression Patterns  
Based Upon 83 Million Basepairs of cDNA Sequence  
Nature 377, 3-174 (1995)  
Other ESTs: EST49388 THC13013  
Contact: Venter, JC  
The Institute for Genomic Research  
932 Clopper Rd, Gathersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423  
Email: tdbinfo@tdb.tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please contact the TIGR Database  
(tdbinfo@tdb.tigr.org)  
Seq primer: M13-21.

FEATURES  
source Location/Qualifiers  
1..263  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):103035"  
/db\_xref="taxon:9606"  
/map="21"  
/clone\_lib="Human Gall bladder"  
/note="Organ: gallbladder"

BASE COUNT 66 a 68 c 49 g 75 t 5 others  
ORIGIN

Query Match 6.7% Score 17; DB 20; Length 263;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 AGAAGTAGAGAGCTGT 69  
|||||  
Db 151 AGAAGTAGAGAGCTGT 167

RESULT 7  
T35391 324 bp mRNA EST 06-SEP-1995  
LOCUS EST83977 Human Parathyroid gland Homo sapiens CDNA 5' end similar  
DEFINITION to None, mRNA sequence.  
ACCESSION T35391  
NID 9617489  
VERSION T35391.1 GI:617489  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 324)  
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,  
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S.,  
Kelley,J.M., Klinek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,  
Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,  
Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,  
Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R.,  
Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,  
Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrite,A.,  
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,  
Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,  
Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wang,J., Xu,C.,  
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,  
Hasselaine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.

Initial Assessment of Human Gene Diversity and Expression Patterns  
Based Upon 83 Million Basepairs of cDNA Sequence  
Nature 377, 3-174 (1995)  
Other ESTs: THC13013  
Contact: Venter, JC  
The Institute for Genomic Research  
932 Clopper Rd, Gathersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423  
Email: tdbinfo@tdb.tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please contact the TIGR Database  
(tdbinfo@tdb.tigr.org)  
Seq primer: M13 Reverse.

FEATURES  
source Location/Qualifiers  
1..324  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):106646"  
/db\_xref="taxon:9606"  
/clone\_lib="Human Parathyroid gland"  
/note="Organ: parathyroid gland"

BASE COUNT 71 a 75 c 63 g 115 t  
ORIGIN

Query Match 6.7% Score 17; DB 20; Length 324;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 53 AGAAGTAGAGAGCTGT 69  
|||||  
Db 198 AGAAGTAGAGAGCTGT 214

RESULT 8  
LOCUS T52594 313 bp mRNA EST 06-FEB-1995  
DEFINITION yb21c03.r1 Stratagene fetal spleen (#937205) Homo sapiens cDNA  
clone IMAGE:71812 5', mRNA sequence.  
ACCESSION T52594  
NID 9654454  
VERSION T52594.1 GI:654454  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 313)  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,  
Chissee, S., Dietrich, N., Dubuque, T., Favello, A., Gish, M.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Merrit, M.  
TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
High quality sequence stops: 170  
Source: IMAGE Consortium, LNL.  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: M13RP1  
High quality sequence stop: 170.  
Location/Qualifiers  
1. .313  
/organism="Homo sapiens"  
/db\_xref="GDB:493477"  
/db\_xref="taxon:9606"  
/map="22"  
/clone="IMAGE:71812"  
/clone\_lib="Stratagene fetal spleen (#937205)"  
/tissue\_type="fetal spleen"  
/dev\_stage="fetal"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: spleen; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
01190 df. Pooled spleens. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCGGAG  
3' -3' adaptor sequence: 5' CTCGACTTTTCTTTTCTTTT 3'."  
BASE COUNT 45 a 86 c 77 g 103 t 2 others  
ORIGIN  
Query Match 6.7%; Score 17; DB 20; Length 313;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 134 TGTCTTTGGGGCCCTTT 150  
|||||  
Db 249 TGTCTTTGGGGCCCTTT 265  
RESULT 9  
LOCUS R09245 318 bp mRNA EST 05-APR-1995  
DEFINITION yf26q03.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
IMAGE:1128020 3' similar to SP:IP9S\_RAT P29995 INOSITOL

ACCESSION R09245  
NID 9761168  
VERSION R09245.1 GI:761168  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 318)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marr, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 783  
High quality sequence stops: 128 Source: IMAGE Consortium, LNL This  
clone is available royalty-free through LNL; contact the IMAGE  
Consortium (info@image.lnl.gov) for further information.  
Insert length: 783 Std Error: 0.00  
Seq primer: -21m13  
High quality sequence stop: 128.  
Location/Qualifiers  
1. .318  
/organism="Homo sapiens"  
/db\_xref="GDB:480181"  
/db\_xref="taxon:9606"  
/map="16"  
/clone="IMAGE:128020"  
/clone\_lib="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pTF3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - Oligo(df) primer  
[5' AACTGAGAGATTATTAAGATCTTTTCTTTTCTTTT 3']  
[5' AACTGAGAGATTATTAAGATCTTTTCTTTTCTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pTF3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."  
BASE COUNT 97 a 77 c 67 g 77 t  
ORIGIN  
Query Match 6.7%; Score 17; DB 21; Length 318;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 106 ATCAAGTTCAGACAAAT 122  
|||||  
Db 22 ATCAAGTTCAGACAAAT 38  
RESULT 10  
LOCUS D52818/c 311 bp mRNA EST 14-SEP-1995  
DEFINITION HOM089B02B Clontech human fetal brain polyA+ mRNA (#6535) Homo  
sapiens cDNA clone GEN-089B02 5', mRNA sequence.  
ACCESSION D52818  
NID 9953054  
VERSION D52818.1 GI:953054  
1,4,5-TRISPHOSPHATE-BINDING PROTEIN TYPE 2 RECEPTOR ;, mRNA  
sequence.  
ACCESSION R09245  
NID 9761168  
VERSION R09245.1 GI:761168  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 311)  
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shimomura,H., Takauchi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H., Shin,S. and Nakamura,Y.  
TITLE Fujiwara et al. (1995)  
JOURNAL Unpublished (1995)  
COMMENT On Nov 29, 1993 this sequence version replaced gi:503005.

CONTACT: Tsutomu Fujiwara  
Otsuka GEN Research Institute  
Otsuka Pharmaceutical Co., Ltd  
463-10 Kagasuno Kawanuchi-cho, Tokushima, Tokushima, 771-01 Japan  
Tel: 0886-65-2888  
Fax: 0886-37-1035.

FEATURES  
source Location/Qualifiers  
1..311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Clontech human fetal brain polyA+ mRNA (#6535)"

BASE COUNT 98 a 58 c 79 g 75 t 1 others

ORIGIN

Query Match 6.7%; Score 17; DB 23; Length 311;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 AGAAGTAGAGGAGCTGT 69  
|||||

Db 139 AGAAGTAGAGGAGCTGT 123

RESULT 11  
L38062 333 bp mRNA EST 07-NOV-1996  
LOCUS BNAF0409E Mustard flower buds Brassica rapa cDNA clone F0409, mRNA  
DEFINITION sequence.  
ACCESSION L38062  
NID 9887237  
VERSION L38062.1 GI:887237  
KEYWORDS EST  
SOURCE Brassica rapa.  
ORGANISM Brassica rapa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 333)  
AUTHORS Lim,C.O., Kim,H.Y., Kim,M.G., Lee,S.I., Chung,W.S., Park,S.H., Hwang,I. and Cho,M.J.  
TITLE Expressed sequence tags of Chinese cabbage flower bud cDNA  
JOURNAL Plant Physiol. 111 (2), 577-588 (1996)  
MEDLINE 96255495  
COMMENT On May 9, 1995 this sequence version replaced gi:803855.

CONTACT: Lim,C.O., Kim,M.G., Hwang,I. and Cho,M.J.  
Plant Molecular Biology and Biotechnology Research Center  
Email: pmbr@nongae.gsnu.ac.kr.

FEATURES  
source Location/Qualifiers  
1..333  
/organism="Brassica rapa"  
/strain="pekinensis"  
/db\_xref="taxon:3711"  
/clone\_lib="F0409"  
/clone\_lib="Mustard flower buds"

BASE COUNT 92 a /note="Devel\_stage = flower bud"  
52 c 77 g 112 t

ORIGIN

Query Match 6.7%; Score 17; DB 23; Length 333;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 TTCAGATCAAGTTCAGA 117  
|||||

Db 36 TTCAGATCAAGTTCAGA 52

RESULT 12  
N41057 365 bp mRNA EST 22-JAN-1996  
LOCUS Y53906.s1 Soares\_multiple\_sclerosis\_2NBHMSP Homo sapiens cDNA  
DEFINITION clone IMAGE:277306 3', mRNA sequence.  
ACCESSION N41057  
NID 9164655  
VERSION N41057.1 GI:1164655  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 365)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maita,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisan,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On May 5, 1995 this sequence version replaced gi:797873.

CONTACT: Wilson RK  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 174  
Source: IMAGE Consortium, LIND  
This clone is available royalty-free through LIND; contact the IMAGE Consortium (info@image.lind.gov) for further information.  
Seq primer: m3 -40 forward  
High quality sequence stop: 174.

FEATURES  
source Location/Qualifiers  
1..365  
/organism="Homo sapiens"  
/db\_xref="GDB:3895682"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:277306"  
/clone\_lib="Soares\_multiple\_sclerosis\_2NBHMSP"  
/sex="male"  
/tissue\_type="multiple sclerosis lesions"  
/dev\_stage="Age 46"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified polylinker V-type; phagemid; Site\_1: Not I - Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis lesions from one patient was kindly provided by Dr. Kevin G. Becker (NINDS/NIH)."

BASE COUNT 75 a 92 c 62 g 129 t 7 others  
ORIGIN

Query Match 6.7%; Score 17; DB 24; Length 365;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 AGAAGTAGAGAGCTGT 69  
|||||  
Db 208 AGAAGTAGAGAGCTGT 224

RESULT 13  
N66420 313 bp mRNA EST 08-MAR-1996  
LOCUS  
DEFINITION y240c05.s1 Morton Fetal Cochlea Homo sapiens CDNA clone  
IMAGE:285512 3', mRNA sequence.

ACCESSION N66420  
NID 91218545  
VERSION N66420.1 GI:1218545  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 313)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

On May 5, 1995 this sequence version replaced gi:797648.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: ml3 -40 forward.  
Location/Qualifiers

## FEATURES

source  
1. 313  
/organism="Homo sapiens"  
/db\_xref="GDB:3892629"  
/db\_xref="taxon:9606"  
/clone="IMAGE:285512"  
/clone\_1lb="Morton Fetal Cochlea"  
/tissue="cochlea"  
/dev\_stage="16-22 week fetus"  
/note="Organ: SOLR cells (kanamycin resistant)"  
/note="Organ: ear; Vector: pBluescript SK-; Site: 1: EcoRI;  
Site: 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned  
unidirectionally. Primer: Oligo dt. Fetal cochlea. normal.  
378 of inserts <0.5 kb; 568 0.5-1.0 kb, 78 >1 kb. Uni-ZAP  
XR Vector. Library constructed by N. Robertson, C. Morton.  
-5' adaptor sequence: 5' GAATTCGCGACAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

## BASE COUNT

69 a 73 c 60 g 111 t

## ORIGIN

Query Match 6.7%; Score 17; DB 25; Length 313;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 AGAAGTAGAGAGCTGT 69  
|||||  
Db 201 AGAAGTAGAGAGCTGT 217

RESULT 14  
N93765 474 bp mRNA EST 20-AUG-1996  
LOCUS  
DEFINITION z664c07.s1 Soares\_fetal\_lung\_NBH19W Homo sapiens CDNA clone  
IMAGE:308364 3', mRNA sequence.

ACCESSION N93765  
NID 91266074  
VERSION N93765.1 GI:1266074  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 474)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

On Oct 18, 1995 this sequence version replaced gi:102689.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 793 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 356.  
Location/Qualifiers

## FEATURES

source  
1. 474  
/organism="Homo sapiens"  
/db\_xref="GDB:125177"  
/db\_xref="taxon:9606"  
/clone="IMAGE:308364"  
/clone\_1lb="Soares\_fetal\_lung\_NBH19W"  
/dev\_stage="19 weeks"  
/note="Organ: lung; Vector: pT73D (ampicillin resistant)"  
/note="Organ: lung; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dt) primer  
[5'-TGTTACCAATCGAATGAGGAGGAGCGCCCAATTTTCTTTTCTTT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Benito  
Soares and M. Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NBH19W."

BASE COUNT 103 a 127 c 78 g 164 t 2 others

## ORIGIN

Query Match 6.7%; Score 17; DB 25; Length 474;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 53 AGAAGTAGAGAGCTGT 69  
|||||

DB 211 AGACTAGAGAGCTGT 227

RESULT 15

LOCUS

DEFINITION

W44889 377 bp mRNA EST 10-OCT-1996  
 zc05g08.r1 Soares.parathyroid\_tumor\_NBHPA Homo sapiens cDNA clone  
 IMAGE:3214705 similar to gb:M28212 RAS-RELATED PROTEIN RAB-6  
 (HUMAN); mRNA sequence.

ACCESSION

NID 91328989

VERSION W44889.1 GI:1328989

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
 human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Eliston, R., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Riklin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.

TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT On Sep 21, 1992 this sequence version replaced gi:279339.

CONTACT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 2660 Std Error: 0.00  
 Seq primer: mob.REGA-ET.

FEATURES

SOURCE

Location/Qualifiers  
 1..377  
 /organism="Homo sapiens"  
 /db\_xref="GDB:1259128"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:321470"  
 /clone.lib="Soares.parathyroid\_tumor\_NBHPA"  
 /tissue\_type="parathyroid tumor"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: parathyroid gland; Vector: pT7T3D  
 (Pharmacia) with a modified polylinker; Site\_1: Not I;  
 Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer  
 15..  
 TGTTCACATCTGAGAGGAGCGCGCCGACCAATTTTTTTTTTTTTTTTTTTT  
 T-3'), double-stranded cDNA was size selected, ligated to  
 Eco RI adapters (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of a modified pT7T3  
 vector (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M.Fatima Bonaudo. RNA from sporadic parathyroid  
 adenomas was kindly provided by Dr. Stephen Marx, National  
 Institute of Diabetes and Digestive and Kidney Diseases,  
 NIH"

BASE COUNT 92 a 109 c 84 g 89 t 3 others

ORIGIN

Query Match 6.7%; Score 17; DB 26; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 23;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GAAC TAGAGAGAGCTGTC 70

DB 257 GAAC TAGAGAGAGCTGTC 273

Search completed: August 6, 1999, 09:22:41  
 Job time: 6232 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:11:12 ; Search time 3264.21 Seconds  
(without alignments)  
281.572 Million cell updates/sec

Title: US-09-049-696-7  
Perfect score: 289  
Sequence: 1 GAATATCCAACTGATGAT.....AGCTTGAGAGTAAAGGATTA 289

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_ph.\*  
7: gb\_ph.\*  
8: gb\_ph.\*  
9: gb\_ph.\*  
10: gb\_ph.\*  
11: gb\_ph.\*  
12: gb\_ph.\*  
13: gb\_ph.\*  
14: gb\_ph.\*  
15: gb\_ph.\*  
16: gb\_ph.\*  
17: gb\_ph.\*  
18: gb\_ph.\*  
19: gb\_ph.\*  
20: gb\_ph.\*  
21: gb\_ph.\*  
22: gb\_ph.\*  
23: gb\_ph.\*  
24: gb\_ph.\*  
25: gb\_ph.\*  
26: gb\_ph.\*  
27: gb\_ph.\*  
28: gb\_ph.\*  
29: gb\_ph.\*  
30: gb\_ph.\*  
31: gb\_ph.\*  
32: gb\_ph.\*  
33: gb\_ph.\*  
34: gb\_ph.\*  
35: gb\_ph.\*  
36: gb\_ph.\*  
37: gb\_ph.\*  
38: gb\_ph.\*  
39: gb\_ph.\*  
40: gb\_ph.\*  
41: gb\_ph.\*  
42: gb\_ph.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result Query  
No. Score Match Length DB ID Description  
1 288 99.7 3311 11 AF039400 Homo sapi

2	288	99.7	2826	42	AF127036	AF127036 Homo sapi
3	200.6	69.4	2937	12	AB017156	AB017156 Mus muscu
4	164.4	56.9	35278	11	AF039401	AF039401 Homo sapi
5	111	38.4	3317	3	AF001261	AF001261 Bos tauru
6	111	38.4	3288	3	AF001262	AF001262 Bos tauru
7	111	38.4	2820	3	AF001263	AF001263 Bos tauru
8	106.2	36.7	3022	12	AF047838	AF047838 Mus muscu
9	106.2	36.7	3471	12	AF052746	AF052746 Mus muscu
10	106.2	36.7	3471	12	AF052746	AF052746 Mus muscu
11	94.2	32.6	3604	9	AB026833	AB026833 Homo sapi
12	84.6	29.3	3415	42	AF043976	AF043976 Homo sapi
13	34.2	11.8	222	8	AF009121	AF009121 Parasitau
14	34	11.8	161667	11	AC005483	AC005483 Homo sapi
15	34	11.8	120408	34	HSJ519P24	HSJ519P24
16	33.6	11.6	35412	7	SPAC48	SPAC48
17	32.6	11.3	75777	9	HS276K20	HS276K20
18	32.4	11.2	1484	1	MHAJ1656	MHAJ1656
19	32.4	11.2	2354	1	MHAJ5054	MHAJ5054
20	32.4	11.2	12905	2	AE000768	AE000768 Agulifex a
21	32.2	11.1	9419	36	AF050740	AF050740 Plasmodiu
22	31.8	11.0	174428	11	AC005821	AC005821 Homo sapi
23	31.6	10.9	2489	4	XIGLUSLSA	XIGLUSLSA
24	31.6	10.9	106179	10	HS790B6	HS790B6
25	31.6	10.9	119562	34	AC006779	AC006779 Caenorhab
26	31.6	10.9	22948	37	CELP13A2	CELP13A2
27	31.4	10.7	197900	11	AC002531	AC002531 Homo sapi
28	31	10.7	124191	11	AC005153	AC005153 Homo sapi
29	30.8	10.7	1952	1	MHAJ5053	MHAJ5053 Mycoplasma
30	30.8	10.7	1802	1	MHP50	MHP50
31	30.8	10.7	133153	34	HSJ799G3	HSJ799G3
32	30.8	10.7	259037	35	AC005911	AC005911 Homo sapi
33	30.6	10.6	8581	8	AF049110	AF049110 Zea mays
34	30.6	10.6	57991	11	AF015593	AF015593 Homo sapi
35	30.4	10.5	12979	1	D90773	D90773 E.coli geno
36	30.4	10.5	12835	2	AE000232	AE000232 Escherich
37	30.4	10.5	662	9	HMPRPB	HMPRPB
38	30.2	10.4	1803	7	MHAJ5055	MHAJ5055 Mycoplasma
39	30.2	10.4	3075	1	YSCF0M12A	YSCF0M12A
40	30.2	10.4	103682	8	SCU12980	SCU12980 Saccharomyc
41	30.2	10.4	128779	9	HS29K1	HS29K1
42	30.2	10.4	146058	11	AC003668	AC003668 Homo sapi
43	30.2	10.4	187543	11	AC005678	AC005678 Homo sapi
44	30.2	10.4	151964	11	AF047825	AF047825 Homo sapi
45	30.2	10.4	159948	34	AC005103	AC005103 Homo sapi

## ALIGNMENTS

RESULT 1  
AF039400 3311 bp mRNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCICL1) mRNA,  
DEFINITION complete cds.  
ACCESSION AF039400  
NID 94009457  
VERSION AF039400.1 GI:4009457  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human CICL1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES  
Source Location/Qualifiers

1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/note="hCLCA1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"

gene

CDS

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 99.7%; Score 288; DB 11; Length 3311;

Best Local Similarity 99.7%; Pred. No. 1.5e-83;

Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAAATATCCAACTGATGATGGAATTTGCTGCTGACGATGGGGAAGACACACTAT 60  
DB 1545 GAAATATCCAACTGATGATGGAATTTGCTGCTGACGATGGGGAAGACACACTAT 1604  
QY 61 AAGTGGTGGCTTTAAGAGAGTCAAAAGTNGTGCATCATCCACAGTCGCTTTGGG 120  
DB 1605 AAGTGGTGGCTTTAAGAGAGTCAAAAGTNGTGCATCATCCACAGTCGCTTTGGG 1664  
QY 121 GCCCTCTGACAGCTCAAGACTAGAGAGAGCTGTCCAAATGACAGAGAGTTTACACACATA 180  
DB 1665 GCCCTCTGACAGCTCAAGACTAGAGAGAGCTGTCCAAATGACAGAGAGTTTACACACATA 1724  
QY 181 TGCTTCAGATCAAGTTTCAGAACATGGCCCTCATTTGCTTTGGGCGCTTTTATCAGG 240  
DB 1725 TGCTTCAGATCAAGTTTCAGAACATGGCCCTCATTTGCTTTGGGCGCTTTTATCAGG 1784  
QY 241 AAATGAGAGCTGTCTCTCAGAGCTCCATCCAGCTTGAAGTAAGGATTA 289  
DB 1785 AAATGAGAGCTGTCTCTCAGAGCTCCATCCAGCTTGAAGTAAGGATTA 1833

RESULT 2  
AF127036 2826 bp mRNA PRI 16-Apr-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CaCC1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2826)

Agnel, M., Verma, T. and Culouscou, J.-M.

Cloning of three human homologs of bovine epithelial chloride channel

Unpublished

2 (bases 1 to 2826)

Agnel, M. and Culouscou, J.-M.

Direct Submission

Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des

Carrières, Neuilly-Malmaison 92500, France

Location/Qualifiers

1. .2826

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/tissue\_type="small intestine; colon"

1. .2826

/gene="CaCC1"

5. .2749

/note="CaCC1"

/note="bovine epithelial chloride channel homolog"

/product="calcium-activated chloride channel protein 1"

/codon\_start=1

/protein\_id="A025487.1"

/db\_xref="PID:94585469"

/db\_xref="GI:4585469"

/translation="MGPKSSVFILHLHLEGLSLSLIQNNNGYEGIVADPNP

EDFTLIQIKDMVTOASLYLEATGKRFYFNVAIIPETWTKADYVRPLETYNA

DVLAESTPPNDPEPTDQNGCEKGERIHLTPDIFAGKLAIEGPGKAFVHEMAH

LRMGVDFEYNDDEKFLYLSNGRIQAVRCAGITGNVVKCOGSCYTRKCFENVTGL

YKGECEFLVOSROTEKASIMFAOHVDSIYECFEDONHNEAPKONOKNLSRWETI

RDSDFKRTTPTOPNPPTFSLQIGORVCIYLDKSGSMATGNRLRLNQGLFL

LQVLEGSWAGMTFDSAAHVOSELIQINSGRDRLARLPAAAGSTISCGLSRA

FTVIRKRPIDGSEIVLTDGEDNTISGCEVEKOSGAIHVTALPSAAQELIELSK

MTGILQTVASDOVONNGLIDAFGLSSGNVAVORSIOLESKGLTQNSOMNNTVIV

DSVAGKDTLEFLITWTPOILMDPSGQKGFVVDKNTKMAVLIQIPILAKGTWKY

SLOASQTLTLVTSRASNAATLPITVTSKNTDTSKPSPLVYVANIROGASPIIRA

SVTALIESVNGKTVTEILDNAGADATDDGYSSRYFTYDNGRSYKVRALGVN

AARRVTPQOSGALYIPGWIENDELQMNPPPELNDQVQKQVCSRTSSGGSFVAS

DVFNAPIDPLPFGQITDLKAEIHGSLINLTWAPGDYDGHGAHKTIRISTSLD

LNRKFNESLQVNTALIPKEANSEVEFLKPEPNTFENGTDLFIALQAVDKVLSKI

SNARVSLFIPQTPPTPPTSPDTSAPCPNIHINSTIPGILHILKIMKWIQELQSLIA

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 99.7%; Score 288; DB 42; Length 2826;

Best Local Similarity 99.7%; Pred. No. 1.5e-83;

Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAAATATCCAACTGATGATGGAATTTGCTGCTGACGATGGGGAAGACACACTAT 60  
DB 1198 GAAATATCCAACTGATGATGGAATTTGCTGCTGACGATGGGGAAGACACACTAT 1257  
QY 61 AAGTGGTGGCTTTAAGAGAGTCAAAAGTNGTGCATCATCCACAGTCGCTTTGGG 120  
DB 1258 AAGTGGTGGCTTTAAGAGAGTCAAAAGTNGTGCATCATCCACAGTCGCTTTGGG 1317  
QY 121 GCCCTCTGACAGCTCAAGACTAGAGAGAGCTGTCCAAATGACAGAGAGTTTACACACATA 180  
DB 1318 GCCCTCTGACAGCTCAAGACTAGAGAGAGCTGTCCAAATGACAGAGAGTTTACACACATA 1377  
QY 181 TGCTTCAGATCAAGTTTCAGAACATGGCCCTCATTTGCTTTGGGCGCTTTTATCAGG 240  
DB 1378 TGCTTCAGATCAAGTTTCAGAACATGGCCCTCATTTGCTTTGGGCGCTTTTATCAGG 1437  
QY 241 AAATGAGAGCTGTCTCTCAGAGCTCCATCCAGCTTGAAGTAAGGATTA 289



	DB	1438	AAATGAGACTGTCTCCTCAAGCGTCCATCATCAGCTTAGAGTAAGGATTATTA	1486
RESULT	3			
AB017156				
LOCUS	Mus musculus	gob-5 mRNA, complete cds.	ROD	04-MAR-1999
DEFINITION	AB017156			
ACCESSION	AB017156			
NID	93721911			
VERSION	AB017156.1	GI:3721911		
KEYWORDS	GOB-5.			
SOURCE	Mus musculus adult intestine goblet cell cDNA to mRNA.			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
REFERENCE	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	Komiya,T., Tanigawa,Y. and Hirohashi,S.			
TITLE	1 (sites)			
JOURNAL	Komiya,T., Tanigawa,Y. and Hirohashi,S.			
MEDLINE	Cloning and identification of the gene gob-5, which is expressed in			
REFERENCE	intestinal goblet cells in mice			
AUTHORS	Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)			
TITLE	Blochm.			
JOURNAL	2 (bases 1 to 2937)			
MEDLINE	Komiya,T., Tanigawa,Y. and Hirohashi,S.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (28-AUG-1998) to the DBJ/EMBL/GenBank databases, Tohru			
TITLE	Komiyu, EBARO, JST, Hirohashi Cell configuration Project, 5-9-9,			
JOURNAL	Tokodai, Tsukuba, Ibaraki 300-2653, Japan			
FEATURES	(E-mail:tkomehccp.jst.go.jp, Tel:81-298-47-7563,			
SOURCE	Fax:81-298-47-5226)			
location/Qualifiers				
1..2937				
/organism="Mus musculus"				
/db_xref="taxon:10090"				
/cell_type="goblet cell"				
/dev_stage="adult"				
/tissue_type="intestine"				
15..2756				
/gene="gob-5"				
15..2756				
/gene="gob-5"				
/codon_start=1				
/protein_id="BAA33743.1"				
/db_xref="pid:d1034712"				
/db_xref="pid:g3721912"				
/db_xref="gi:3721912"				
/translation="MESLRSVFFLLIHLLEGVLSESLIQLNNGYEGIVAIADHDVP				
EEDELIOHKIDMVTQASPYLEAEGRKRYFNVAIIIPESKAPRETRPLETKENNA				
DVALSTSPLENDVEPIHEIGACGEKRIILHPDFLAGKKLTIGQODRTFVEMAH				
FRMCVFENYNDEKEFYLSKGKPOAVRCSAITGKNVRGCGSGITNGCKVIDRVVG				
LKYNCVPEWPHONERKASIMENONINSVCFCEKHNDAAPDONRCULBRTWEY				
IODEBKOTTPMTAQPPAFPSLSLIQGVTCVIVLDSSGMDDRLRNNOARSLE				
LQGVFEKSWGWATFDPSAAYVSSELQLNSGARDDLIRKLPTVASAGVISCSGLPR				
AFVIYIKRRYPVDGEIYLVDGEDNTISSCDLVQSGAILHTVALGPAAKELEOLS				
KMEGLDTSDSQVNNGNLDAFAALSSGNAAIOHSIOLESREVNLNONGMSVYI				
VDSWGDTLETTLWTTHPPIFIWDPSGYVONEFLIDTTKVYLVPGKAYGFMKK				
YSIOASSQTLLIVTSRAASTLPPIYTVPVNKTEKRPSPVYVASISRGASPILR				
ASYVALIESVNGKTYTVELDNGAGADATKMDVYSRFPLFDANGRSYATNALGGY				
TSDDQRAPPNRMARYIDGWLEDGERNPMPREPSTYQDQLFSRTSSGSFVATNR				
VPAAPAPIPDLEPCQITDLDEKASIOGOMLYNLTWAPGDYDHGASNYIIMSTSIVYD				
LRFHFTSLQVNTGLIPIPKASSEIEEFELGCVTFNGTDIFIAIDVKSINKSEIL				
SNIRVSEVFIPAOPPIPEDSTPCPDISINSTIPGIHVLIKIMKMVLGEMQVTLGLH				
BASE COUNT	860 a	718 c	693 g	666 t
ORIGIN				
Query Match	69.4%; Score 200.6; DB 12; Length 2937;			
Best Local Similarity	80.9%; Pred. No. 4.6e-55;			
Matches 233; Conservative	0; Mismatches 55; Indels 0; Gaps 0;			
yy	1 GAATATCCAACTGATGATCTGAATATTGCTGCTGCATCAGCATGGGGAACACACTAT 60			
db	1211 GAAATGATCAACTGATGATCTGAATATTGCTGCTGCATCAGCATGGGGAACACACTAT 1270			

OY	61	AAATGGGCGCTTTAACGAGGCTCAAAACAAGATNGTCATTCATTCACACAGTCGCTTTGGG	120
Db	1271	TAGCAGCTGCTTTGACCTGGTGAAGCAGAGGGGGCCATCATCATCAAGTGGCCCTGGG	1330
OY	121	GCCCTCTCAGCTCAGAACTAGAGAGAGCTCTCCAAAATGACAGAGATTACAGACATA	180
Db	1331	ACGGGCTCCGGCTAAAGAGCTTGGACAGCTTCCAAAATGACAGAGGCCCTGGACACATA	1390
OY	181	TGCTTCAGATCAAGTCTTCAGAAACAATGGCCATCATTTGATGCTTTTGGGCCCTTTCAGG	240
Db	1391	CTCTTCGGATCAGGTTTCAGAACAAATGGTCTTGTGATGCTTTCCAGACAGCTCTCTCAGG	1450
OY	241	AAATGAGACTGTCTCTCAGCCGCTCATTCACACTTGAAGTAAAGGATT	288
Db	1451	AAATGGCGGATCGCTCAGACTCCATTCACACTTGAGAGCAGGGAGAT	1498
RESULT	4		
AF039401		AF039401	35278 bp DNA PRI 15-DEC-1998
LOCUS		Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,	
DEFINITION		complete cds.	
ACCESSION		AF039401	
NID		g4009459	
VERSION		AF039401.1	GI:4009459
KEYWORDS		human.	
SOURCE		human.	
ORGANISM		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
		Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE		1 (bases 1 to 35278)	
AUTHORS		Gruber,A.D., Eblle,R.C., Ji,H.L., Schreur,K.D., Fuller,C.M. and	
TITLE		Pauli,B.U.	
		Genomic cloning, molecular characterization, and functional	
		analysis of human CLCA1, the first human member of the family of	
		Ca2+-activated Cl- channel proteins	
JOURNAL		Genomics 54 (2), 200-214 (1998)	
MEDLINE		99047526	
REFERENCE		2 (bases 1 to 35278)	
AUTHORS		Gruber,A.D., Eblle,R. and Pauli,B.U.	
TITLE		Direct Submission	
JOURNAL		Submitted (19-DEC-1997) Department of Pathology, College of	
FEATURES		Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA	
source		Location/Qualifiers	
		1..35278	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="1"	
		/map="1p22-p31"	
		1..1617	
		/gene="hCLCA1"	
		1..33522	
		/gene="hCLCA1"	
		767..1265	
		/rpt_family="L1 transposable element"	
		/rpt_type="dispersed	
		1587..1596	
		/gene="hCLCA1"	
		/note="putative"	
		1618..1708	
		/gene="hCLCA1"	
		1..number=1	
		join(1618..1708,1962..2383,6667..6807,6979..7126,	
		9656..9801,15454..15631,18591..18809,19783..20010,	
		22253..22427,24522..24628,26644..26859,27447..27708,	
		28766..28936,31840..32079,32919..33522)	
		/gene="hCLCA1"	
		/product="calcium-dependent chloride channel-1"	
		1962..2383	
		/gene="hCLCA1"	
		1..number=2	
		join(2222..2383,6667..6807,6979..7126,9656..9801,	
		15454..15631,18591..18809,19783..20010,22253..22427	
CDS			

```
24522..24628,26644..26859,27447..27708,28766..28936,
31840..32079,32919..33310)
/gene="hclca1"
/codon_start=1
/product="calcium-dependent chloride channel-1"
/protein_id="AAC95429.1"
/db_xref="pid:g4009460"
/db_xref="gi:409460"
/translation="MPPKSVFIIILHLLLEGALNSLIQLNNNGEIVVAIDPNV
EEDTLIQIKDMVQASLYLEFATGRREFKNVALIPEWTMKADYRPLETKNA
DLVAESTPPDEDEYTEQMGNGCEGSEHILTPFIIAGKLAEGPQGFVEMAH
LMGVEDEYNDEKFEYLSNGRQIAVSGIGTVNKKCGSGCYTRCPENKTKY
YKSGEFLVOSROTEKASIMFQHDVSIYECTEONHNRKANONOKNCRSTREVI
RDSDFPKTTPHTPTPPNPTFSLDIOGRIYCVLDKSGSMATGNRLNKLNOAGFL
LOTVEIGSVWGVTFDSAAHVQSELIQINSQSDRDTLAKRLPAASGGTSCSGLRSA
FVIRKKYPTDSEIVLLDGEDNTISGCFNWKQSGAIIHVALGPSAAOELIELSK
MTGGLQIVASDOVQNNGLIDAFGALSNGAVSORSIOLESGLTQSNOMNGTVIV
DSYVGDITFLITTTOPQIILMDPSGGKGFVDNRTKAYLOIGIAKVTWKY
SLQASSOTLTLTVTSRASNALPITVTSKTKTKRSPSPPLVYANIRQASPTLRA
SVTALIESVNGKTVLELIDNGADATKDDSVYRITTTIDTNGRISVKRALGVN
ARRRVIPQSGALTIIPQWENDEIQMNPPEIRKDVQVHQCFSFTSSGSEVAS
DVNPAPIDPLEPPGQITDLNAIEHGSLLNLTWTPGDDYHGHAKYIIRISITILD
LNDKNESLQVNTALIPKANESEVFLFKPENITFENGTDLFIQAVDVKLSEI
SNIRVSLFIPQTPPETPSPDETSAPCPNHINSTIPGIIHLKIMMWIGELQISA
"
6667..6807
/gene="hclca1"
/number=3
6979..7126
/gene="hclca1"
/number=4
9696..9801
/gene="hclca1"
/number=5
13434..15631
/gene="hclca1"
/number=6
18591..18809
/gene="hclca1"
/number=7
19783..20010
/gene="hclca1"
/number=8
22253..22427
/gene="hclca1"
/number=9
24522..24628
/gene="hclca1"
/number=10
26644..26859
/gene="hclca1"
/number=11
27447..27708
/gene="hclca1"
/number=12
28766..28936
/gene="hclca1"
/number=13
31840..32079
/gene="hclca1"
/number=14
32919..33322
/gene="hclca1"
/number=15
BASE COUNT 11245 a 6870 c 7230 g 9933 t
ORIGIN
```

```
Query Match 56.9%; Score 164.4; DB 11; Length 35278;
Best Local Similarity 98.8%; Pred. No. 3.9e-43;
Matches 165; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
1 GAAATATCCAACTGATGATCTGAATATGTGCTGACGAGTGGGGAACAACACTAT 60
```

```
|||||
Db 22264 GAAATATCCAACTGATGATCTGAATATGTGCTGACGAGTGGGGAACAACACTAT 22323
Oy 61 AGTGGGCTTTTACGAGGCAACAAGTNGGCCATCATCCACAGCTGCTTTGGG 120
Db 22324 AGTGGGCTTTTACGAGGCAACAAGTNGGCCATCATCCACAGCTGCTTTGGG 22383
Oy 121 GCCCTCTCAGCTCAGACTAGAGAGAGCTGTCCAAATGACAGAG 167
Db 22384 GCCCTCTCAGCTCAGACTAGAGAGAGCTGTCCAAATGACAGAGT 22430
```

```
RESULT 5
AF001261
LOCUS
DEFINITION
Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA,
complete cds.
ACCESSION
AF001261
VERSION
g2623762
AF001261.1 GI:2623762
KEYWORDS
SOURCE
ORGANISM
Bos taurus.
Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae;
Bovinae; Bos.
REFERENCE
1 (bases 1 to 3317)
Elble, R.C., Widom, J., Gruber, A.D., Abdel-Chany, M., Levine, R.,
Goodwin, A. and Pauli, B.U.
Cloning and characterization of Lu-ECAM-1 suggest it is an
endothelial chloride channel
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 3317)
Elble, R.C., Widom, J., Gruber, A.D., Abdel-Chany, M., Levine, R.,
Goodwin, A. and Pauli, B.U.
Direct Submission
TITLE
Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,
Ithaca, NY 14853, USA
FEATURES
Location/Qualifiers
1..3317
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_type="aortic endothelial cells"
/clone="1"
/tissue_type="lung"
63..2780
/note="endothelial adhesion molecule; chloride channel"
/codon_start=1
/product="Lu-ECAM-1"
/protein_id="BAB86529.1"
/db_xref="PID:g2623763"
/db_xref="GI:2623763"
/translation="MYCLNVLILFLILHLLPGKSMVNLINNGDYIYAIPNPSVE
DEKLINIKEMVTEASTYLFHATKRVRYSIILIPMTKSKSEYFLIPKOSYQAD
VIVANPYLKYGDPPYTIQYRGCEKGYHHPNLLINPMTKSKSEYFLIPKOSYQAD
RNGIDERYVDPPYISKRNTIATRCSHITGIVNVRKCGGSCITSLCRDSOTG
LVBACQFLPKKSQTAKEIMFPLSHVTECTEKTHTNTEPNLNQNMCKSTWDY
INNSVDFQNTSPMTENPPTHPTFELSKSKQVOCVLVDKSGSMABEDRLQMNQAE
LYLQIYIEKSGSLVGWTFEDVAEIONHLIRITDNYQIKITAKLPQVANGTSGRL
KAGFOAIHSDQSTGSEIILTDGEDNEINCFEDVRSQAIHITIALGPSAAELE
TISNMTGGYREFRANKDITGLTNAFSRISRSISGSIQOALQLESKAKITGKRNVGTY
PVDSTVGNDFPVVVTWITQKPEIVYQDPRGKRYKISDRCDRLNRSARLQIPGIAET
GWTTSYSLNNHSSQMLTYVTVTRARSPTIPPIVATAMSOHTAHYPSPMIVYQAVSQ
GFLPVGLISVIAIILETDGHOVTLLELWNGAKRDYKNDGIYSRFTDYINGRSLK
YHAQARNNTARLNLRQPNKVLVYGVYENGKRIINPDPREKVDLAKAKIEDFSRLT
SGSFTVSGAPPPGPNHPSVFPSPKITLDEAKFKEDYIOLSMTPAGNVLDKGRANSYII
RISKSPMDROEDFNATLVNTSNLIPKFAKSGKNEPFRPHRVNGTKFTIYQAIN
EANLISEVSHYQAIKFIPLPDSYHDKTISLITLAILGLPMIFSIF"
```

```
BASE COUNT 1084 a 676 c 634 g 923 t
ORIGIN
```

Query Match

38.4%; Score 111; DB 3; Length 3317;

PVDSTVGNDFEFVVTWTIQPEIVLODPKGGKKTSDFEKEDKLNIRSAQLQIGIAET  
GTWTYSLNNHASSQMLTVTVTTRASPPTIPPVIATAHMSQHTAHYPSPMIVYAQVSQ

1043 a 707 c 645 g 893 t  
VHAQARNNNTARLNRLRPQNKVLVYGVENGCI ILNPPRPEVKDDLAKIKIDFSRLT  
SGGSFETVSGAPPPGNHPSVFPSPKITPLEAKFKEDYIQLSWTAAPGNVLDKGXS"

Query Match: 11; Score: 38.48; DB: 3; Length: 3286;

Matches 181; Conservative 0; Mismatches 91; Indels 6; Gaps 1;  
11 ACTGATGCATCTAATATGTCCTGCTGACGGATGGGGAGACACACATATACTGGGTC 70  
QY

DB	12/4	ACTTCTGGTTCGAAATCATACTATTAACTGATGAGGGGAGATTAATGAATAAATTCACGC	1333
QY	71	TTTAAACGAGGTCAACAAAGTNTGCCATCATCCACACAGTGGCTTTGGGGCCCTTCGA	130

DB 1334 TTGAGGATGTAAACGAGTGGTGCATTCACACACCATTTGCTGAGACCCTCTGCT 1393

QY 131 GCTCAGAACTAGAGGAGCTGTCCAAATAGACAGAGAGTTTACAGACATATGCTTCAGAT 190

Db 1394 GCCAAGCACTGGAGACATGTCTCAATATAGACAGGAGATATC-----GTTTTTTGCC 144 /

QY 191 CAA GTTCAGACAATAGCCCTCATTTGATGCTTTTGGGGGCCCTTTCATCAGGAATGAGACT 250

DB 1448 AATTAAGCATACTAGCCCTTACTAATGCTTCAGTAGAATTTCACTAGAAAGTGAAGC 150/  
QY 251 GGTCTCTAGCGCTCCATCCAGCTTGAGAGTAAGGATT 288

Db	1508	ATCACTCAGCAGCGCTATTCAGITGGAAAGCAAGCCIT	1545
1			

AF001263					
LOCUS	AF001263	2820 bp	mRNA	19-NOV-1997	
DEFINITION	Bos taurus clone 3 endothelial adhesion molecule				
				Lu-ECAM-1 mRNA,	

AF001263.1  
AF001263  
92623766  
92623766  
GI:2623766

SOURCE  
BOS TAURUS.  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Mammalia; Artiodactyla; Ruminantia; Bovora; Bovidae.  
Bovidae.

REFERENCE  
Bovinae; BOS.  
1 (bases 1 to 2820)  
Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Geddes, J., and Paul, D. H.

JOURNAL  
NUMBER  
3 (March)  
45 (1970)

TITLE  
Cloning and characterization of  
endothelial chloride channel

It is suggested that the

TITLE	AUTHORS	YEAR	DOI
Direct Submission	Goodwin, A. and Pauli, B. U.	1997	10.1002/1522-2675(199709)19:3<221::AID-HLCA221>3.0.CO;2-1
Submitted (25-Nov-1997)	Gruber, A. D., Abdel-Ghany, M., Levine, R., Eddle, R. C., Widdow, J. J.	1997	10.1002/1522-2675(199709)19:3<221::AID-HLCA221>3.0.CO;2-1
Accepted (25-Nov-1997)	Gruber, A. D., Abdel-Ghany, M., Levine, R., Eddle, R. C., Widdow, J. J.	1997	10.1002/1522-2675(199709)19:3<221::AID-HLCA221>3.0.CO;2-1
Published (25-Nov-1997)	Gruber, A. D., Abdel-Ghany, M., Levine, R., Eddle, R. C., Widdow, J. J.	1997	10.1002/1522-2675(199709)19:3<221::AID-HLCA221>3.0.CO;2-1
Revised (25-Nov-1997)	Gruber, A. D., Abdel-Ghany, M., Levine, R., Eddle, R. C., Widdow, J. J.	1997	10.1002/1522-2675(199709)19:3<221::AID-HLCA221>3.0.CO;2-1
Final (25-Nov-1997)	Gruber, A. D., Abdel-Ghany, M., Levine, R., Eddle, R. C., Widdow, J. J.	1997	10.1002/1522-2675(199709)19:3<221::AID-HLCA221>3.0.CO;2-1

FEATURES	SOURCE
11naca, N1 14835, USA	Location/Qualifiers
1. .2820	
/formatism="BOS taurinet"	

```

/od_xc1c="LAXON:3513"
/clone="3"
/tissue_type="lung"
/cell_type="epithelial cells"

```

```

CDS
154..2008
    /note="endothelial adhesion molecule; chloride channel"
    /codon_start=1
    /product="E1u-ECAM-1"

```

```
/proc/sys-id=4006032.1  
/db_xref="PID:g2623767"
```

/db\_xref="GI:2623767"  
/translation="MWLCNVLFLTLHLPLGKSSMNVNINNGYDGIIVAINPSVE  
DEKLINIKEMVTEASTYLEFHAKRRVRYFNVSILIPMTKSKSEYFIIPKQESYDQAD  
VIVANPLYKDDPYTLOYGRGCKGKIHTPELNLNHNHIGSRVGRVEMHML  
RMGJFEYNDOPFYISKRNTIENRCSHTITGIVNFKKPGGSCITISCRRSORG  
LYEAKCPFKPSQTAKESTIMFMSISYEFCEKTHNTAPMLQNMCKGKSTMY  
IMNSVDFONTSPTEMPPTHTSLSKSRVCLVDKSGMSAEERLQNMQAL  
LYLIOVLEKSLVGMVFDVAELIQLNHLTRITDNYOKITAKLPVANGSTICREL  
KAGFOAIHSDQSTSGSEIILLTDGENEINISCEPDYKRSQAIHTITAGPSAKELE  
TUSNMGTGREFRANKDITGLTNAPSRISRSSTIOQAIOLSKALKITGRKNGVY  
PVDSTGNDPEFVVTWITQKEIYLODPKGRKYSDFEKDKLNRSLRLOIPGIAET  
GVTYSLNNHSSOMLTWYTRASPTIPPTATAMSOHTAHYSPMVAVYQO  
GELPVLGIVYAIITETEDGHQVITELMDNGKRDYKNDGITSRTFDYNGVYSIK  
VHAQRNNTARLNRPONKALYVGVENKILINPREPVKODLAKAIEDFSRLT  
SGSEFTVSGAPPGNHPSPVPPSKITPLEAKFKEDYIOLSWTAGNVLDRKASGSP  
MSRSHVAKLELOHOSFO"

Query Match 38.4%; Score 111; DB 3; Length 2820;  
Best Local Similarity 65.1%; Pred. No. 7.3e-26;  
Matches 181; Conservative 0; Mismatches 91; Indels 6; Gaps 1;

OY 11 ACTGATGATCTGAATGTGCTGCTGACGGATGGGGAAGCAACACATATAGTGGTGC 70  
Db 1409 ACTCTGTTCTGAATATATCTATTACTGATGGGAAGATGAATAATTTCATGCG 1468  
OY 71 TTAAAGAGTCTCAACAAAGTNGTCCATCCACACAGTCCGCTTGGGCGCTCTGCA 130  
Db 1469 TTGAGAGTGTAAAGAAAGAGTGTGCATCCACACATGCTGCTGGACCCCTGCT 1528  
OY 131 GCTCAGAGCTAGAGAGCTGCTCCAAATGACAGAGGTTTACAGACATATGCTTCAGAT 190  
Db 1529 GCCAAGAGCTGAGACATCTGCAATATGACAGAGAGATATC-----GTTTCTTCC 1582  
OY 191 CAAGTTCAGAACATNGGCTCATTTGCTTTGGGCGCTTTCATCAGAAATGAGACT 250  
Db 1583 AATAAGACATATGCGCTTCTATGCTTTCATGATAGAAATTTTGAAGTGAAGC 1642  
OY 251 GTCTCTCAGCGCTCCATCCAGCTTGAAGTGAAGGATT 288  
Db 1643 ATCACTCAGCAGGCTATTCAGTTGGAAGCAAGGCTT 1680

RESULT 8  
LOCUS BTU36445 2984 bp mRNA MAM 09-FEB-1996  
DEFINITION Bos taurus calcium-activated chloride channel mRNA, complete cds.  
ACCESSION U36445  
NID g1184065  
VERSION U36445.1 GI:1184065  
KEYWORDS  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 2984)  
AUTHORS Cunningham, S.A., Awayda, M.S., Bubljen, J.K., Ismailov, I.I.,  
Arrate, M.P., Berdiev, B.K., Benos, D.J., and Fuller, C.M.  
TITLE Cloning of an epithelial chloride channel from bovine trachea  
JOURNAL U. Biol. Chem. 270 (52), 31016-31026 (1995)  
MEDLINE 96125078  
REFERENCE 2 (bases 1 to 2984)  
AUTHORS Benos, D.J.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,  
University of Alabama at Birmingham, BHSB 706, Birmingham, AL  
35294, USA

FEATURES  
source 1. 2984  
Location/Qualifiers  
\* /organism="Bos taurus"

/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19. 2730  
/note="epithelial chloride channel"

/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="PID:g1184066"  
/db\_xref="GI:1184066"

/translation="MWPLVILFLTLHLPLGKSSMNVNINNGYDGIIVAINPSVE  
DEKLINIKEMVTEASTYLEFHAKRRVRYFNVSILIPMTKSKSEYFIIPKQESYDQAD  
VIVANPLYKDDPYTLOYGRGCKGKIHTPELNLNHNHIGSRVGRVEMHML  
RMGJFEYNDOPFYISKRNTIENRCSHTITGIVNFKKPGGSCITISCRRSORG  
LYEAKCPFKPSQTAKESTIMFMSISYEFCEKTHNTAPMLQNMCKGKSTMY  
IMNSVDFONTSPTEMPPTHTSLSKSRVCLVDKSGMSAEERLQNMQAL  
LYLIOVLEKSLVGMVFDVAELIQLNHLTRITDNYOKITAKLPVANGSTICREL  
KAGFOAIHSDQSTSGSEIILLTDGENEINISCEPDYKRSQAIHTITAGPSAKELE  
TUSNMGTGREFRANKDITGLTNAPSRISRSSTIOQAIOLSKALKITGRKNGVY  
PVDSTGNDPEFVVTWITQKEIYLODPKGRKYSDFEKDKLNRSLRLOIPGIAET  
GVTYSLNNHSSOMLTWYTRASPTIPPTATAMSOHTAHYSPMVAVYQO  
GELPVLGIVYAIITETEDGHQVITELMDNGKRDYKNDGITSRTFDYNGVYSIK  
VHAQRNNTARLNRPONKALYVGVENKILINPREPVKODLAKAIEDFSRLT  
SGSEFTVSGAPPGNHPSPVPPSKITPLEAKFKEDYIOLSWTAGNVLDRKASGSP  
MSRSHVAKLELOHOSFO"

BASE COUNT 1027 a 583 c 561 g 813 t  
ORIGIN

Query Match 36.7%; Score 106.2; DB 3; Length 2984;  
Best Local Similarity 64.0%; Pred. No. 2.7e-24;  
Matches 178; Conservative 0; Mismatches 94; Indels 6; Gaps 1;

OY 11 ACTGATGATCTGAATGTGCTGCTGACGGATGGGGAAGCAACACATATAGTGGTGC 70  
Db 1231 ACTCTGTTCTGAATATATCTATTACTGATGGGAAGATGAATAATTTCATGCG 1230  
OY 71 TTAAAGAGTCTCAACAAAGTNGTCCATCCACACAGTCCGCTTGGGCGCTCTGCA 130  
Db 1291 ATTGAGAGTGTAAAGAAAGAGTGTGCATCCACACATGCTGCTGGACCTTGTCT 1350  
OY 131 GCTCAGAGCTAGAGAGCTGCTCCAAATGACAGAGGTTTACAGACATATGCTTCAGAT 190  
Db 1351 GCCAAGAGCTGAGACATCTGCAATATGACAGAGAGATATC-----C 1404  
OY 191 CAAGTTCAGAACATNGGCTCATTTGCTTTGGGCGCTTTCATCAGAAATGAGACT 250  
Db 1405 AATAAGACATATGCGCTTCTATGCTTTCATGATAGAAATTTTGAAGTGAAGC 1464  
OY 251 GTCTCTCAGCGCTCCATCCAGCTTGAAGTGAAGGATT 288  
Db 1465 ATCACTCAGCAGACTATTCAGTTGGAAGCAAGGCTT 1502

RESULT 9  
LOCUS AF047838 3022 bp mRNA ROD 26-NOV-1998  
DEFINITION Mus musculus calcium-sensitive chloride conductance protein-1  
(mClCA1) mRNA, complete cds.  
ACCESSION AF047838  
NID g3925280  
VERSION AF047838.1 GI:3925280  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3022)  
AUTHORS Gandhi, R., Elble, R.C., Gruber, A.D., Schreier, K.D., Ji, H.L.,  
Fuller, C.M., and Pauli, B.U.  
TITLE Molecular and functional characterization of a calcium-sensitive  
chloride channel from mouse lung



QY 251 GTCCTACGCGCTCCATCCAGCTTGAGAGTAAAGGATT 288  
DB 1493 GTCCTACGAGGCTCTGAGTTGGAGAGCAAGCCTT 1530

## RESULT 11

AB026833 3604 bp mRNA PRI 26-MAY-1999  
LOCUS Homo sapiens mRNA for chloride channel protein, complete cds.  
DEFINITION AB026833  
ACCESSION AB026833  
NID 94887600  
VERSION AB026833.1 GI:4887600  
KEYWORDS chloride channel protein.  
SOURCE Homo sapiens adult corneal epithelium cDNA to mRNA,  
clone\_11b:lambda ZAP clone:lambda 13802.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Itoh, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.  
TITLE Isolation and characterization of chloride channel in human corneal  
epithelium  
JOURNAL Unpublished (1999)  
REFERENCE 2 (bases 1 to 3604)  
AUTHORS Itoh, R., Kawamoto, S. and Okubo, K.  
TITLE Direct Submission  
JOURNAL Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases. Relko  
Itoh, Institute for Molecular and Cellular Biology, Osaka  
University, Yamada-oka 1-3, Suita, Osaka 565-0871, Japan  
(E-mail:relko@imcb.osaka-u.ac.jp, Tel:81-6-6879-7992,  
Fax:81-6-6877-1922)

FEATURES  
source

1. 3604  
/location="Qualifiers"  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone\_11b="lambda 13802"  
/clone\_11b="lambda ZAP"  
/dev\_stage="adult"  
/map="1p32"  
/tissue\_type="corneal epithelium"  
42. 2873  
/codon\_start=1  
/product="chloride channel protein"  
/protein\_id="BA077810.1"  
/db\_xref="PID:41041578"  
/db\_xref="PID:94887601"  
/db\_xref="GI:4887601"  
/translation="WTORSIAGPICNLEKFTLVLAALSEPFLGAGVOLQDNGYNGIL  
IAINPQVPEONLISNIKEMITEASFYLFNATKRKVFERNKILIPATWKAANNKIK  
QSEYKAVNIYTDVGAHGDPTLYRGCGKEGYIHFTPNFLINDLITAGNSRGR  
VEFEHMAHLRMGVEDEYINDKPEYINGNOIKVRGSSDITGIFVCEGPGPOENCIT  
SKLEFGCTFIYNSONATASIMPMOSTSYVEECNASTHQAEPNOCNOCSTRSM  
DVITDSADPHHSFRMNGTELPPEPFFSVQGDYKVCILVDYSSMAEADLLDLOQA  
AEFLIMQVEIHTFVGASFDSEKELRAQLHINSNDKRLVSLPTVYSAKDISI  
CSGLKGFVEVERKNGAYGSMILVSGDKRLNCLPTVLSGSSITHSIALGSSA  
PNLELSRLTGKLFYVDIGNSNMDIASGSDIGDFOHIOLESTENKPPH  
QKNTVVDNTVNDIMFLVWQASGPEIILFDPGGRKYTNFTNLTPRTSLMT  
PQTAQGHWTYTLNNTHTSLQALKYTVTSRASNSAVPATEAFERDLSHFPPIVNI  
YANKOGFPIPLNTATVATPEPMDPTLLDLDGAGADYIKNDGYSRTFFSFAN  
GRSLKHAVNNSPISITPAHSIPGSHATVGYTANGIONNARKSKVGRNEERKNG  
FSRVSSGGSFVSLVAPAGHPDVPCKIIDLAVKYEEBELTSLWAGDEPDGQAT  
SEYIRMSKSLIONIDDENNALIVNTSKNPQAGIREFITFSPOISTNGEPHONGET  
HESHRIYVAILRAMDRNSLOSASVNIQAOLPIPNSDPVPARDYLLIKGLVTLANGLLG  
ITCLIVYTHHTLSRKRKDKKENGKTKLL"

BASE COUNT 1113 a 740 c 732 g 1017 t 2 others

ORIGIN

Query Match 32.6%; Score 94.2; DB 9; Length 3604;  
Best Local Similarity 58.7%; Pred. No. 2.3e-20;  
Matches 162; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 10 AACTGATGATCTGAATTTGCTGCTGACGGATGGGAGAACACATATAGTGGGG 69  
DB 1268 AGTTATGGCTCTGATGATATATTAGTACCGAGAGATGATGCTTGGCAATTG 1327  
QY 70 CTTTAAAGAGCTCAACAAAGTNGTCCATTCACACAGTGGCTTGGGCCCTGCG 129  
DB 1328 CTTACCACTGCTGCTGACGAGTGGTTCACAAATTCATCTCATTCGCTGGGTTCATGCG 1387  
QY 130 AGCTCAAGAACTAGGAGGAGCTGCCAAATGACAGAGGTTAGACATATGCTTACA 189  
DB 1388 AGCCCAATTCGAGAGATTTATCACGCTTTACAGAGAGTTTAAAGTCTTGTTCACA 1447  
QY 190 TCAAGTTCAGAACAAATGGCCCTCATTTGATGCTTTGGGCCCTTCATCAGAAATGAGC 249  
DB 1448 TATATCAAACTCCATATGATGATGATGCTTCAAGTAAATTCCTGTGAACTGAGGA 1507  
QY 250 TGTCTCTACGCGCTCCATCCAGCTTGAGAGTAAAGG 285  
DB 1508 CATTTCCAGCAACATATTCACGCTTGAAAGTACAG 1543

## RESULT 12

AF043976 3415 bp mRNA PRI 07-APR-1999  
LOCUS AF043976  
DEFINITION Homo sapiens CLCA homolog (hCLCA3) mRNA, complete cds.  
ACCESSION AF043976  
NID 94572288  
VERSION AF043976.1 GI:4572288  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 3415)  
AUTHORS Gruber, A.D. and Pauli, B.U.  
TITLE Molecular cloning and biochemical characterization of a truncated,  
secreted member of the human family of Ca<sup>2+</sup>-activated Cl<sup>-</sup> channels  
JOURNAL Blochm. Biophys. Acta 1444 (3), 418-423 (1999)  
MEDLINE 99196715  
REFERENCE 2 (bases 1 to 3415)  
AUTHORS Gruber, A.D., Eblle, R.C. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

FEATURES  
source

1. 3415  
/location="Qualifiers"  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/gene="hCLCA3"  
/gene="hCLCA3"  
19. 807  
/gene="hCLCA3"  
/codon\_start=1  
/product="CLCA homolog"  
/protein\_id="AA023734.1"  
/db\_xref="PID:94572289"  
/db\_xref="GI:4572289"  
/translation="MFSLSKVLFLSLSPVKSLSVLTNNNGDGIYVIAINPSVE  
DEVLIONIKEMYTEASTHLFHAIKORAYRNVSIILPMYKSKSEYLIPKOTYQAD  
VYADLYLKYGDPTLYQGGCKGQYIHFTPNFLINDLITAGNSRGRVEFGWAHL  
RMGVEDEYVNDQPFYISRNTEATRCSTRIVYVWLNCKASCARFPRDSOTGL  
YEAKCFIRKSTAKESIVPMONDSVTEFECTHNKEANL"

BASE COUNT 1149 a 625 c 630 g 1011 t

ORIGIN

Query Match 29.3%; Score 84.6; DB 42; Length 3415;  
Best Local Similarity 71.2%; Pred. No. 3e-17;  
Matches 111; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
QY 11 ACGATGATGATCTGAATTTGCTGCTGACGAGTGGGAGAACACATATAGTGGGTCG 70  
DB 1228 ACCTTGCTCTGATGATATATCTATTAACAGATGAGGAGGAGATTATCAATAAGCTTATGC 1287









***This Page Blank (uspto)***



JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES Location/Qualifiers

source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="pid:94009458"  
/db\_xref="gi:4009458"  
/translation="MGPKSSVFIIILHLEGALSNSLIQLNNNGEIVVAIDPNP  
EDETLIQIKDMVTOASLYLEATGRKRFYFNVAIIPETWTKADYVRPLEYKNA  
DVLVAESTPPGNDPEYTEMONGCEKGERIHLPPFIAGKLAEGPOGAFFVEMAH  
LRMGVEDEYNDEKFIYLSNGRIQAVRCSAGITGVVKKCOGSGCYTRKCFNVTGL  
YKGECEVILSROTERKASIMFAOHVDSIVFECTEONNKNRKNOKCKLBSTEWI  
RDSDFRKTTPMTTOPPNTPTSLQIQRIYCLVLDKSGSMATGNRLNNOGOLF  
LQVELGSMGVNTPDVAHVQSELIOINSQSDRDITAKRLPAASGSGTSCGLNSA  
FTVIRKKYPTDGSSEIVLTDGEDNTISGFNEVKOSGAIHTVALGSAOELIELSK  
MTGGLQTVASDOVONNGLIDAFGLSNGNAVORSIOLESGLTIQNSOMNGVIV  
DSYAGKDTFLITWTTOPQIILMPDSGKGGEVVDKRTKMAVLOIPGAKGTWY  
SLQASQTLILVTSTRASNATLPITVTSKNTKDSKRPSPPLVYVANIROGASPIIRA  
SVTALESVNGKTVLELIDNGAGADARKDGYVSRTTITDNRKRYKRAIAGVNA  
RARRVITPQSGALITPGWIEDEIOMNPPREINKDDVQKQVCFSTSSGGSFVAS  
DVNRAPIDLPPEQITDLKAEIHGSLINLWTAPDDYDHTAHKYLIRISTSLD  
LRKFNESLQVNTALIPKRNSEVFLFKENITFENGDTLFAIOAVDVKLSBI  
SNIRAVSLFIPQTPPETPSDETSAFCPNHINSTIPGIIHLKIMKMWIGELQLSIA

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 252; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 5.2e-63;  
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAAGATTGTGTGTTAGTCTTGAACAATCTGGAAGCATGGCAGCTGTACCGCCTC 60  
DB 1261 CAAAGATTGTGTGTTAGTCTTGAACAATCTGGAAGCATGGCAGCTGTACCGCCTC 1320  
OY 61 AATGACTGAATCAAGAGGCGCAGCTTTCTGTCGAGACAGTTGAGGTGGGTCCTGG 120  
DB 1321 AATGACTGAATCAAGAGGCGCAGCTTTCTGTCGAGACAGTTGAGGTGGGTCCTGG 1380  
OY 121 GTTGGATGTTGACATTTGACAGTCTGCCCATTGTACAAAGTGAAGTCAATACAGATAAC 180  
DB 1381 GTTGGATGTTGACATTTGACAGTCTGCCCATTGTACAAAGTGAAGTCAATACAGATAAC 1440  
OY 181 AGTGCAGTGAACAGGAGACACACTGCCCAAAAGATTACTGTCGACAGCTTCAGAGGAGC 240  
DB 1441 AGTGCAGTGAACAGGAGACACACTGCCCAAAAGATTACTGTCGACAGCTTCAGAGGAGC 1500  
OY 241 TCCATCTGCAGC 252  
DB 1501 TCCATCTGCAGC 1512

RESULT 2  
AF127036 2826 bp mRNA PRI 16-Apr-1999  
LOCUS  
DEFINITION Homo sapiens calcium-activated chloride channel protein 1 (CaCC1)  
ACCESSION AF127036  
MID 94585468  
VERSION AF127036.1 GI:4585468

KEYWORDS  
SOURCE  
ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
1 (bases 1 to 2826)  
Agnel, M., Vermaat, T. and Culouscou, J.-M.  
Cloning of three human homologs of bovine epithelial chloride  
channel

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
DIRECT SUBMISSION  
Carrières, Rueil-Malmaison 92500, France

FEATURES  
SOURCE

1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="CaCC1"  
5. .2749  
/note="CaCC1"  
/gene="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AA025487.1"  
/db\_xref="pid:94585469"  
/db\_xref="gi:4585469"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 252; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 5.2e-63;  
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAAAGATTGTGTGTTAGTCTTGAACAATCTGGAAGCATGGCAGCTGTACCGCCTC 60  
DB 914 CAAAGATTGTGTGTTAGTCTTGAACAATCTGGAAGCATGGCAGCTGTACCGCCTC 973  
OY 61 AATGACTGAATCAAGAGGCGCAGCTTTCTGTCGAGACAGTTGAGGTGGGTCCTGG 120  
DB 974 AATGACTGAATCAAGAGGCGCAGCTTTCTGTCGAGACAGTTGAGGTGGGTCCTGG 1033  
OY 121 GTTGGATGTTGACATTTGACAGTCTGCCCATTGTACAAAGTGAAGTCAATACAGATAAC 180  
DB 1034 GTTGGATGTTGACATTTGACAGTCTGCCCATTGTACAAAGTGAAGTCAATACAGATAAC 1093  
OY 181 AGTGCAGTGAACAGGAGACACACTGCCCAAAAGATTACTGTCGACAGCTTCAGAGGAGC 240  
DB 1094 AGTGCAGTGAACAGGAGACACACTGCCCAAAAGATTACTGTCGACAGCTTCAGAGGAGC 1153  
OY 241 TCCATCTGCAGC 252  
DB 1153 TCCATCTGCAGC 1153



	RESULT	4							
	LOCUS	ABO17156	2937 bp	mRNA	ROD	04-MAR-1999			
	DEFINITION	Mus musculus gob-5 mRNA, complete cds.							
	ACCESSION	ABO17156							
	NID	G3721911							
	VERSION	ABO17156.1	GI:3721911						
	KEYWORDS	Gob-5.							
	SOURCE	Mus musculus adult intestine goblet cell cDNA to mRNA.							
	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;							
	REFERENCE	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
	AUTHORS	Komiyama,T., Tanigawa,Y. and Hirohashi,S.							
	TITLE	Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice							
	JOURNAL	Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)							
	REFFLINE	99160866							
	REFERENCE	2 (bases 1 to 2937)							
	AUTHORS	Komiyama,T., Tanigawa,Y. and Hirohashi,S.							
	TITLE	Direct Submission							
	JOURNAL	Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases, Tohru Komiyama, EPRG, JST, Hirohashi Cell Configuration Project, S-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan (E-mail:tkomihc@jst.go.jp, Tel:81-298-47-7563, Fax:81-298-47-5226)							
	FEATURES	Location/Qualifiers							
	source	1..2937							
		/organism="Mus musculus"							
		/db_xref="taxon:10090"							
		/cell_type="goblet cell"							
		/dev_stage="adult"							
		/tissue_type="intestine"							
		15..2756							
		/gene="gob-5"							
		15..2756							
		/codon_start=1							
		/protein_id="BAA33743.1"							
		/db_xref="PID:d1034712"							
		/db_xref="PID:g3721912"							
		/db_xref="GI:3721912"							
		/translation="MESLSPVFLIIHLLEGSESLIINNNGEGIVAIINDHPV EDALIHQIMVTQASPYLEFATGRKRIENVAILLIPESKAKPETRPRELFETKNA DYLVSTISLGNDPEYIEHIGAGCGEKGRIRHLDPDLAAGKLITGYEQDRFEVENAH FRMVCVEYNNDDEKFYLSSKGPOAVNRSAALTGNVRCOGSGCINGKCVIDRWVG LYKNCFVPNDEPHONERKASIMEFNONSIVECTEKHNODAPDNORCLRETWVN IOESEDFKQITPMACPAPPAFYLSLIQGVICVLDSGSMLNDDRLNNRNOSRLF ILDTVEOGSWMGVTFPSAAVOSELKONGADRDLLIKHLPTVSAGGTSCGGTLT AFNYTKRRYPDGSEIYLTDGEENTISSCDLYKOGSAILHYVALPAAKELEQLS KMTEGLDIOTSSDOYQNNGLVDFAPLSSGNALIOHSIQLESRYVNLQNOMNKGSYT VVDSVGDVLTFLITWIHTPIFIWDPSGEQNF ILDTTTKVALVDPGTAKGFVKFK YSVIASSQTLLITVTSRAASATLPPIIVTPVNNKTGPFSPVVASIRGASPILR ASTVALIESVNGKVTVTELNDGAGADATKDGVSRFTFPANGRVSYKIMALGCV TSDSORAAPENRMYIDGNTIEDGEEVNNRPRTSVYODOLCFERTSSGGSFVANTN VPAAPIPDELPPPCOIDLKASIOGMNVMTWAPDGDHGRASVYTIIMNSTIYD LRHEINTSLQVNTGLIPKESASEITEFELGNTPEENGCDIFIALDVKSNSKSI SNIRKRSVFTRADEPPIPEDSTPPCPDISINSTIPGILHVLIKMKWLGEKQVTLGLH"							
	BASE COUNT	860 a 718 c 693 g 666 t							
	ORIGIN								
	Query Match	65.2%	Score 164.2;	DB: 12;	Length 2937;				
	Best Local Similarity	78.7%	Pred. No. 1.2e-37;						
	Matches 196;	Conservative 0;	Mismatches 53;	Indels 0;	Gaps 0;				
Oy	1	CAGAAGATTGCTGTTTGTCCTTACAACAATCTGGAAGCATGGCGACTGTAACGCCTC	60						
Dd	927	CAGAAGANTGTGCTGTTAGTCTTGATAGAAGCCGGAGACATGCTGAAGAAGATGCTT	986						
Oy	61	ATGCACTTAATCAAGACGCCACGCTTTCTCTGCTGACAGACAGTTGAGCTGGGCTCTGG	120						

QY	121	GTGGGATGGTACATTTGACAGTGTGCTGCCCATGTACAAAGTGAACCTCTACAGATTAAC	180
Db	1047	GTCCGGGATGGTACCTTTGACAGTGTGCTGCTGTGTGTAACAAAGGAACCTCAAAAGTTAAAC	1106
QY	181	AGTGACAGTACAGAGGACACACACTCCGCCAAAATATTACCTGCAGCAGACTTCAGAGAGGACG	240
Db	1107	AGTGATGCTGCAGACAGATCTGCTGATTCAGACACTTACCCACAGATATCTCAGAGAGGACA	1166
QY	241	TCCATCTCG	249
Db	1167	TCTATATGC	1175
RESULT	5		
LOCUS	AF047838		
DEFINITION	AF047838	3022 bp mRNA	ROD
ACCESSION	AF047838	Mus musculus calcium-sensitive chloride conductance protein-1 (mclcal1) mRNA, complete cds.	26-NOV-1998
VERSION	AF047838.1	GI:3925280	
KEYWORDS			
SOURCE		house mouse.	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		(bases 1 to 3022)	
TITLE		Gandhi, R., Eble, R.C., Gruber, A.D., Schneur, K.D., Ji, H.L., Fuller, C.W. and Paul, B.U.	
JOURNAL		Molecular and functional characterization of a calcium-sensitive chloride channel from mouse lung	
MEDLINE		J. Biol. Chem. 273 (48), 32096-32101 (1998)	
AUTHORS		2 (bases 1 to 3022)	
TITLE		Gandhi, R., Eble, R.C., Gruber, A.D. and Paul, B.U.	
JOURNAL		Direct Submission	
FEATURES		Submitted (12-FEB-1998) Vet. Pathology, Cornell Univ., WMC C4-146, Ithaca, NY 14856, USA	
source		Location/Qualifiers	
gene		1. .3022	
CDS		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/tissue_type="whole Lung"	
		1. .3022	
		/gene="mclcal1"	
		18. .2726	
		/gene="mclcal1"	
		/note="calcium-activated chloride channel; Alternative chloride channel to CFTR that may influence severity of cystic fibrosis in different tissues; mclac-1"	
		/codon_start=1	
		/product="calcium-sensitive chloride conductance protein-1"	
		/protein_id="AAC79982.1"	
		/db_xref="pid:g3925281"	
		/db_xref="GI:3925281"	
		/translation="MPEGLQVLLFLTHLLQNTSSMVHLNSNGYEGVIAINSPVEPDERLPSIKEMVQAATYLEFASGGRVFNISILIVMTKSKPEYILMPKESYDKADVIAVPHLQGDDEPYLQYGCGRQGIYFTPEFLTDNLRIYGPGRVYEMAHUHLRMGVEDYNDOPFMYRKNTIEKTRSTRTGNVNVNBERGNCVTRACROSKTLLYEKCTFIPKTIQTAGSIMPMONLNSVEFCEKNNANPEPNLOKMKRNRSDVYIKTSADRONAPMRKTEAPPPPTSLKSRKRVCLVDKSGSDKEDRLIRNQAELIYLOIVERKSMVGLVTFEDSAHLQNTYIKIKITSSDYOKITANLPQASGGTISGQLAGCAQTISDSQSTSGSEIVLLTGDGEDNIGKTSFEAVSRSAIHTTIALGSAALEFTVLSMTQGLIRFANKDLSLIDAFSRISSTGSYSOALQLESKFAVRAGAWINGTVYLDSTVGDTFELITVMYTKRPEIILQDKGKRTSDPODKLNRSAILOIPGAETGCTWYISITGCTGSEIVTMYTTRARSPTEPILLATRAHOSQTAQVSRMIVYARVSGPFLPVGANTALILEAEHGHQVTLLELNDNGAGDYKANDCITRTFTDYGNGRYSLKVRVQAQRNTRLSLRQNRKSLYTPGVYENKIVLANPRPVQEALEATVEDFRVTSGGSGFTVSGAPPPDDHARVPEPSKYVTLDEAFIDGYIHLTWTAQKVLNDRAHNYIIRMSQ	

Query Match	43.3%	Score 109;	DB 12;	Length 3022;
Best Local Similarity	65.3%	Pred. No. 9.9e-22;		
Matches 160;	Conservative 0;	Mismatches 85;	Indels 0;	Gaps 0;
Db	937	a	710 c	661 g 728 t
Query	5	GAATTGTTGTTTAACTCCTTGACAAATCTGGAAGCATGGCGACTGTAAACCGCTCAATC	64	
Db	937	GAGTGTGTGCTGGTGTGATTAATCTGGAAGCATGGACAAAGAACCGCTTATTTC	996	
Query	65	GACCTAATCAACGACGCCAGCTTTTCTTCGCTGCAACACACTGAGCTGGGCTCGGTTG	124	
Db	997	GATTAATCAACGACGCCAGCTTTTCTTCGCTGCAACACACTGAGCTGGGCTCGGTTG	1056	
Query	125	GGATGTGTCATTTGACAGTGTGCTGGCCATGTGTCAAGGAAGTCTACAGATTAACAGT	184	
Db	1057	GATTGTGTCATTTGACAGTGTGCTGGCCATGTGTCAAGGAAGTCTACAGATTAACAGT	1116	
Query	185	GCAGTGACAGGAGCAGCTCTGGCCAAAGATTACTGTGACAGACTTCAGAGGAGCTCA	244	
Db	1117	GTAGTGACTACCAAAAGATACACCGCAAACTCCCAACAGAGCTCTGTGTGAATTCAA	1176	
Query	245	TCTGC 249		
Db	1177	TTTGC 1181		
RESULT 6				
LOCUS	AF052746	3471 bp	MOD	08-APR-1999
DEFINITION	Mus musculus chloride channel cacc mRNA, complete cds.			
ACCESSION	AF052746			
NID	93560546			
VERSION	AF052746.1	GI:3560546		
KEYWORDS				
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euthelia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 3471) Romio, L., Musante, L., Cinfl, R., Serl, M., Moran, O., Zegarra-Moran, O. and Galietta, L.J.V.			
TITLE	Characterization of a murine gene homologous to the bovine Cacc chloride channel			
JOURNAL	Gene 228 (1-2), 181-188 (1999)			
MEDLINE	99173882			
REFERENCE	2 (bases 1 to 3471) Romio, L., Musante, L., Cinfl, R., Moran, O., Serl, M. and Galietta, L.J.V.			
AUTHORS	Direct Submission Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto Giannina Gaslini, Genova I-16148, Italy			
TITLE	Location/Qualifiers			
JOURNAL	1..3471			
FEATURES				
source	/organism="Mus musculus" /db_xref="taxon:10090" /db_xref="dbest:AA245677" 47..2752 /codon_start=1 /product="chloride channel cacc" /protein_id="AAC35003.1" /db_xref="PID:93560547" /db_xref="GI:3560547" /translation="MVPGLVLEFLTLHLQNTSSMVLNSNGYEGVIAINSPSE DERLIPSIKEMVQASTYLEASQGRVYFIRNLSIVPMKSKREYLPMPRESYRAD VIVADPHLQHGDDPYTLOYGGCGRGQYIHFTPNFLTDNLRIYGPGRVFEHML RMGVEDYNDOPFYMRSKNTIETRTCTGTGNNVYHNEBRCVYCARADRSKRL YERKCTIPDKIQTAGASIMFMQNLNVSEETCKNNNAEPNLQNKCKRRSTWDYI KTSADCFNAPMRGTETAPPTETSLKSRRVVCLVLDKSGSDMKEDRLIRMQAAEL			
CDS				

```

BASE COUNT      1096 a 777 c 736 g 862 t
ORIGIN
Query Match      43.3%; Score 109; DB 12; Length 3471;
Best Local Similarity 65.3%; Pred. No. 9.9e-22;
Matches 160; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

OY 5 GAATGTGTGTGTTTACCTGTTGACAAATCTGGAAAGCATGGCAGCTGTAACCGCTCAATC 64
Db 966 GAGTGTGTGCTGCTGCTGCTGTAATCTGGAAGCATGGACAAAGACCGCTTATATTC 1025
OY 65 GACTGAATCAAGCAGCAGCGCAGCTTTTCCCTGTCGACAGCAGTTGAGCTGGGGCTCTGGATTG 124
Db 1026 GAATCAATCAAGCAGCAGCAACTGTACTTAATCAATCTGTGGAAGAAAGAGCTATAGCTTG 1085
OY 125 GCATGTGATCATTTTGACAGTCTGCTGCCATGTGACAAAGTGAACCTCATACAGATTAACAGTGTG 184
Db 1086 GATTAGTCACATTTTGACAGCCTGCTGCCATCAATCAATTAATTAATAAAATAACGAGTA 1145
OY 185 GCATGGAACAGGAGACACATCCGCCAAAATATTAACCTGCAGCAGCTTCAGAGAGGAGAGCTCCA 244
Db 1146 GTAGAGTACTACCAAAAGATCAACCGCAAAACCTCCCAACAGAGCTTGTGTGGAACCTTAA 1205
OY 245 TCTGC 249
Db 1206 TTTGC 1210

RESULT 7
BTU36445 2984 bp mRNA MAM 09-FEB-1996
LOCUS BTU36445 2984 bp mRNA MAM 09-FEB-1996
DEFINITION Bos taurus calcium-activated chloride channel mRNA, complete cds.
ACCESSION U36445
NID g1184065
VERSION U36445.1 GI:1184065
KEYWORDS
SOURCE
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovine; Bos.
1 (bases 1 to 2984)
Cunningham,S.A., Awayda,M.S., Buben,J.K., Ismailov,I.I.,
Arrate,M.P., Berdiev,B.K., Benos,D.J. and Fuller,C.M.
Cloning of an epithelial chloride channel from bovine trachea
J. Biol. Chem. 270 (52), 31016-31026 (1995)
96125078
2 (bases 1 to 2984)
Benos,D.J.
Direct Submission
Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,
University of Alabama at Birmingham, BHSB 706, Birmingham, AL
35294, USA
location/Qualifiers
1..2984
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_type="epithelial cells"
/tissue_type="trachea"
19..2730
/note="epithelial chloride channel"
/codon_start=1
FEATURES
Source
CDS

```

/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="pid:91184066"  
/db\_xref="gi:1184066"  
/translation="MPEPLVILFLTLHLPGKMSMNVNINNGYDGIYAINPSVE  
DEKLIQNKEMVETASTYLPHATKRRYFRNVSILIPMTKSKSEYILMPQESDQAE  
VIVANPYKHKGGDDDTLOYGRGKGOVIEHTPNLNLNLYIGSRGAFVEMAL  
RMGIDFENQOPFVYSIRNTIETATGCTHTGNNVVKOGGSGITRPRBDOTL  
YEACFTEIPKSTAKESIMFQSHSTECTKTNVAPNPNKMGKSKMVDY  
MNSIDFQNTSPTEKMPOTFESILSKQVCLVDKSGSSSEDRTRMNAAL  
FLIOIIEKGLVGVWTFDVSALIRNNLTKITDDVYENITANLPQEAAGSELT  
AGFOAIQSQOSTSGSEIILITDESDNEHSCIEVQSGVITHTIALPGSAELET  
LSDMGHGRFANDINGLTNFSRISRSISGSIQOITLESKALATTEKWNATVP  
VDSITGNDTFVWMTIKKPELLLODPKRGKTSDEKDLNHSARLRIPJAETG  
TWTSILNHNHSPQILTYTVTRARSPPTPVTAHNSONTAHPSVYIYAGVSG  
FLPVGINVTAIETEDGHVYTLMDGACADYKNDGYSRTFYDGRGSRSLV  
HAEARNNTARLSLROPQNKALYIPQYIENGKILINPPPEVKDILAKEDSERLS  
GGSFTVSGAPGNHPSVLPNKKIIDLKPKFEDHJOLSWTPANVLKGRANSYIIN  
SKSEFLDKQFDNATLVNTSSLKPKKAGSDENFEKPEPFIENGNTFYIYVQAINEA  
NITSFVSNIAAKRIEIPMEDSVPLGTRKISAINLAIPALMILISIV"

BASE COUNT 1027 a 583 c 561 g 813 t

ORIGIN

Query Match 43.1%; Score 108.6; DB 3; Length 2984;

Best Local Similarity 64.5%; Pred. No. 1.3e-21;

Matches 162; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 1 CAAGAATTGTGTTTACTTCTTGACAAATCTGGAAGCATGGCGACTGTAAACGGCTC 60  
DB 934 CAGCGAGTAGTGTGTTGTTGATTAATCTGGAAGCATGCTTCAAGATCGCTC 993  
QY 61 AATGACTGATCAAGAGCGCAGCTTCTCTGACAGCATGGCTGAGCTGCTCTG 120  
DB 994 TTTCGATGATCAAGAGCGCAGCTTCTCTGACAGCATGGCTGAGCTGCTCTG 1053  
QY 121 GTTGGATGTGACATTTGACAGTGTGCTCCATGTACAAAGTACATCAGATAAC 180  
DB 1054 GTTGGATGTGACATTTGACAGTGTGCTCCATGTACAAAGTACATCAGATAAC 1113  
QY 181 AGTGGCATGACAGGACACACTGCCAAAGATTACCTGACAGCTGACAGGAGC 240  
DB 1114 GATGATGATCTTTATGAAATATCAGCAATCTGCTCAAGAGCTAATGTGAGACT 1173  
QY 241 TCCATCTGCAG 251  
DB 1174 TCAATTTGAG 1184

RESULT 8  
AF043976 3415 bp mRNA PRI 07-APR-1999  
LOCUS AF043976 Homo sapiens CLCA homolog (hCLCA3) mRNA, complete cds.  
DEFINITION AF043976 94572288  
NID AF043976.1 GI:4572288  
VERSION AF043976.1  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 3415)  
AUTHORS Gruber,A.D., and Pauli,B.U.  
TITLE Molecular cloning and biochemical characterization of a truncated,  
secreted member of the human family of Ca2+-activated Cl- channels  
JOURNAL Biochim. Biophys. Acta 1444 (3), 418-423 (1999)  
MEDLINE 99196715  
REFERENCE 2 (bases 1 to 3415)  
AUTHORS Gruber,A.D., Elble,R.C., and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers

source 1. 3415  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene 1. 3415  
/gene="hCLCA3"  
CDS 19. 807  
/gene="hCLCA3"  
/codon\_start=1  
/product="CLCA homolog"  
/protein\_id="AAD23734.1"  
/db\_xref="pid:94572289"  
/db\_xref="gi:4572289"  
/translation="MPEPLVILFLTLHLPGKMSMNVNINNGYDGIYAINPSVE  
DEKLIQNKEMVETASTYLPHATKRRYFRNVSILIPMTKSKSEYILMPQESDQAE  
VIVANPYKHKGGDDDTLOYGRGKGOVIEHTPNLNLNLYIGSRGAFVEMAL  
RMGIDFENQOPFVYSIRNTIETATGCTHTGNNVVKOGGSGITRPRBDOTL  
YEACFTEIPKSTAKESIMFQSHSTECTKTNVAPNPNKMGKSKMVDY  
MNSIDFQNTSPTEKMPOTFESILSKQVCLVDKSGSSSEDRTRMNAAL  
FLIOIIEKGLVGVWTFDVSALIRNNLTKITDDVYENITANLPQEAAGSELT  
AGFOAIQSQOSTSGSEIILITDESDNEHSCIEVQSGVITHTIALPGSAELET  
LSDMGHGRFANDINGLTNFSRISRSISGSIQOITLESKALATTEKWNATVP  
VDSITGNDTFVWMTIKKPELLLODPKRGKTSDEKDLNHSARLRIPJAETG  
TWTSILNHNHSPQILTYTVTRARSPPTPVTAHNSONTAHPSVYIYAGVSG  
FLPVGINVTAIETEDGHVYTLMDGACADYKNDGYSRTFYDGRGSRSLV  
HAEARNNTARLSLROPQNKALYIPQYIENGKILINPPPEVKDILAKEDSERLS  
GGSFTVSGAPGNHPSVLPNKKIIDLKPKFEDHJOLSWTPANVLKGRANSYIIN  
SKSEFLDKQFDNATLVNTSSLKPKKAGSDENFEKPEPFIENGNTFYIYVQAINEA  
NITSFVSNIAAKRIEIPMEDSVPLGTRKISAINLAIPALMILISIV"

BASE COUNT 1149 a 625 c 630 g 1011 t

ORIGIN

Query Match 41.8%; Score 105.4; DB 42; Length 3415;

Best Local Similarity 63.7%; Pred. No. 1.1e-20;

Matches 160; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1 CAAGAATTGTGTTTACTTCTTGACAAATCTGGAAGCATGGCGACTGTAAACGGCTC 60  
DB 931 CAGCGTAGTGTGTTGTTGATTAATCTGGAAGCATGCTTCAAGATCGCTC 990  
QY 61 AATGACTGATCAAGAGCGCAGCTTCTCTGACAGCATGGCTGAGCTGCTCTG 120  
DB 991 TTTCGATGATCAAGAGCGCAGCTTCTCTGACAGCATGGCTGAGCTGCTCTG 1050  
QY 121 GTTGGATGTGACATTTGACAGTGTGCTCCATGTACAAAGTACATCAGATAAC 180  
DB 1051 GTTGGATGTGACATTTGACAGTGTGCTCCATGTACAAAGTACATCAGATAAC 1110  
QY 181 AGTGGCATGACAGGACACACTGCCAAAGATTACCTGACAGCTGACAGGAGC 240  
DB 1111 GATGATGATCTTTATGAAATATCAGCAATCTGCTCAAGAGCTAATGTGAGACT 1170  
QY 241 TCCATCTGCAG 251  
DB 1171 TCAATTTGAG 1181

RESULT 9  
AF001261 3317 bp mRNA MM 19-NOV-1997  
LOCUS AF001261 Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION AF001261 complete cds.  
ACCESSION AF001261  
NID 92623762  
VERSION AF001261.1 GI:2623762  
KEYWORDS Bos taurus.  
SOURCE Bos taurus.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 3317)  
AUTHORS Elble,R.C., Widom,J., Gruber,A.D., Abdel-Ghany,M., Levine,R.,  
Goodwin,A., and Pauli,B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
endothelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3317)  
AUTHORS Elble,R.C., Widom,J., Gruber,A.D., Abdel-Ghany,M., Levine,R.,  
Goodwin,A., and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers



source

1.3317

/organism="Bos taurus"

/db\_xref="taxon:9913"

/cell\_type="aortic endothelial cells"

/clone="1"

/tissue\_type="lung"

63..22780

/note="endothelial adhesion molecule; chloride channel"

/codon\_start=1

/product="Lu-ECAM-1"

/protein\_id="AAB86529.1"

/db\_xref="pid:92623763"

/db\_xref="gi:2623763"

/translation="WVLCNLVLEFLTLHLPGMKSSVNLINGIDGIVIAINPSVE  
DEKLENIKEVTEASTYLEFHATKRRVFRNVSILIPMTWKSSEYFIPKOSYQAD  
VIVANPLYKGDDEPTLYOGRGCEKGYIHFPNPLINNHFIYSGRGRVHEVAHL  
RMGIDEXVNDQPEYISRNKTIEATRCGSHITGINVFRKCGGSCITSLGRDSOTG  
LYEACFTLPKRSOTAKESIMPMPSLHSETECTEHTNTEAPNLONMCKGSKWDY  
IMNSVDFONTSPMTENPPTHTPESLKSQVYCVLYDKSGMSAEERLPMONAAE  
LYLQVIERGSLVGMVTFDSVAEIONHLRITDENVYKRTAKLPQVANGSTICRGL  
KAGFOAIIHSDQSTGSEIILLTDEDEINSCFEDVRSQAIIHTIALGSAAELE  
TLSNMTGGRREFANDITGLTNARISRSRSGSITQQAIOLESKALKITGRKRVGTV  
PDSVNGDTEFFVWTIOKPEIYLODPKGRKYSDEKEDLNRSARLQIPGIAET  
GTMWYSLNNHASSOMLVYVTTTARSPTIPPIVATAMHSOHTAHYPSPMIYVAQVSO  
GFLPYGIVSVAIIITEDEGHVYTLMDNGARQDYKNDGITSRYFTDYNGRSLK  
VHAQRNNATLNLKQPOKVLVYPGYENKIIINPREPEYKDLAKAKIEDSERLT  
SGGSVTSVGAAPPNGHPSVFPSPKITTLEAKFKEDYIOLSWTARQENVLDKGRS  
RISKSFROEDFENATLVNTSNLIPKESGSENEFEFPEHFRANGVNGKRFYISVOAIN  
EANLSEVSHIYVOAKFIPLPDSVHDGTRKISEITLAILGLPMIFSVF"

BASE COUNT 1084 a 676 c 634 g 923 t

ORIGIN

Query Match 41.2%; Score 103.8; DB 3; Length 3317;  
Best Local Similarity 63.3%; Pred. No. 3.1e-20;  
Matches 159; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1 CAAGAATTGTGTGTTTAACTCCTTGACAAATCTGACAGCATGGCCACTGTAAACGCCCTC 60  
DB 981 CAGCGGGTAGTCTGTTGTTGATTAATCTGGAAGCATGTCTGCAGAAAGCCGCTCTC 1040

QY 61 AATGCATGAATCAAGCAGCCAGCTTTCCGCTGCTGACAGAGTGGAGCTGCTCG 120  
DB 1041 TTTCAAATTAATCAAGCAGCAGACATTAATCTGATTAATTAAGGAGATCTTAA 1100

QY 121 GTTGGATGATGATTTGACAGTGTGCTCCATGTACAAAGTGAATCATACAGATAAC 180  
DB 1101 GTTGGATGATGATTTGACAGTGTGCTCCATGTACAAAGTGAATCATACAGATAAC 1160

QY 181 AGTGCAGTGACAGGACACACTCGCCAAAGATTACTGACAGCAGCTTCAGAGGAGC 240  
DB 1161 GATGATTAATGTTTACCAAAAGATCACCGCAAAAGCTCTCAAGTACTAATGTTGAG 1220

QY 241 TCCATCTGCAG 251  
DB 1221 TCAATTTGTAG 1231

RESULT 10  
AF001262 3288 bp mRNA MAM 19-NOV-1997  
LOCUS Bos taurus clone 2 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.  
ACCESSION AF001262  
NID 92623764  
VERSION AF001262.1 GI:2623764  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 3288)

AUTHORS Elble,R.C., Widom,J., Gruber,A.D., Abdel-Ghany,M., Levine,R.,  
Goodwin,A. and Paull,B.O.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
JOURNAL endothelial chloride channel  
REFERENCE 2 (bases 1 to 3288)  
AUTHORS Elble,R.C., Widom,J., Gruber,A.D., Abdel-Ghany,M., Levine,R.,  
Goodwin,A. and Paull,B.O.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., WVC C4-146,  
Ithaca, NY 14853, USA  
FEATURES  
source location/Qualifiers  
1..3288  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="2"  
/tissue\_type="lung"  
59..2443  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86530.1"  
/db\_xref="pid:92623765"  
/db\_xref="gi:2623765"

BASE COUNT 1043 a 707 c 645 g 893 t

ORIGIN

Query Match 41.2%; Score 103.8; DB 3; Length 3288;  
Best Local Similarity 63.3%; Pred. No. 3.1e-20;  
Matches 159; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1 CAAGAATTGTGTGTTTAACTCCTTGACAAATCTGACAGCATGGCCACTGTAAACGCCCTC 60  
DB 977 CAGCGGGTAGTCTGTTGTTGATTAATCTGGAAGCATGTCTGCAGAAAGCCGCTCTC 1036

QY 61 AATGCATGAATCAAGCAGCCAGCTTTCCGCTGCTGACAGAGTGGAGCTGCTCG 120  
DB 1037 TTTCAAATTAATCAAGCAGCAGACATTAATCTGATTAATTAAGGAGATCTTAA 1096

QY 121 GTTGGATGATGATTTGACAGTGTGCTCCATGTACAAAGTGAATCATACAGATAAC 180  
DB 1097 GTTGGATGATGATTTGACAGTGTGCTCCATGTACAAAGTGAATCATACAGATAAC 1156

QY 181 AGTGCAGTGACAGGACACACTCGCCAAAGATTACTGACAGCAGCTTCAGAGGAGC 240  
DB 1157 GATGATTAATGTTTACCAAAAGATCACCGCAAAAGCTCTCAAGTACTAATGTTGAG 1216

QY 241 TCCATCTGCAG 251  
DB 1217 TCAATTTGTAG 1227

RESULT 11  
AF001263 2820 bp mRNA MAM 19-NOV-1997  
LOCUS Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.





```
repeat_region /note="Alu" repeat: matches 298. .1 of consensus"
19378. .19673 /note="Alu" repeat: matches 298. .3 of consensus"
repeat_region 19746. .20048 /note="Alu" repeat: matches 1. .301 of consensus"
22313. .22872 /note="match: 5' EST R33748 clone 136236; Paired with EST
R33636 matching this clone; match: 3' EST R33636 clone
136236; Paired with EST R33748 matching this clone;
similar to Alu and L1"
22594. .22727 /note="MIR repeat: matches 146. .1 of consensus"
22774. .23076 /note="Alu" repeat: matches 1. .302 of consensus"
23077. .23176 /note="Alu" repeat: matches 133. .232 of consensus;
incomplete repeat"
23367. .23448 /note="MIR repeat: matches 144. .65 of consensus"
24577. .24929 /note="match: 266939 STS containing (CA) repeat"
24681. .24704 /note="match: 47 copies of GT 100% conserved; differs from
266939"
prim_transcript 25490. .26297 /note="match: multiple ESTs; match: AA234670 AA044231
AA069087 AA043988; match: AA235330 AA044041"
26323. .26622 /note="Alu" repeat: matches 2. .301 of consensus"
26915. .27209 /note="Alu" repeat: matches 1. .295 of consensus"
27990. .28388 /note="match: 5' EST AA463302 clone 796916"
28471. .>30274 /note="match: multiple ESTs; match: W38743 AA340888
AA369572 Z42874 F05173; match: F21915 N42511 N24514 H29366
N92614 AA223147; match: N29066 R19712 AA223148 AA463206
N29819; match: F01420 Z39009"
30818. .30937 /note="MIR repeat: matches 243. .106 of consensus"
31245. .31385 /note="MIR repeat: matches 16. .164 of consensus"
31907. .32206 /note="Alu" repeat: matches 303. .3 of consensus"
32445. .32503 /note="MIR repeat: matches 291. .1 of consensus"
32624. .32914 /note="match: 5' EST AA045796 clone 488037; Paired with
EST AA045797 matching this clone; match: 3' EST AA045797
clone 488037; Paired with EST AA045796 matching this
clone"
35182. .35480 /note="Alu" repeat: matches 1. .299 of consensus"
36160. .36229 /note="Alu" repeat: matches 299. .230 of consensus;
incomplete repeat"
36862. .37111 /note="Alu" repeat: matches 241. .2 of consensus;
incomplete repeat"
37153. .37443 /note="Alu" repeat: matches 1. .302 of consensus"
37465. .37561 /note="MIR repeat: matches 163. .262 of consensus"
38171. .38293 /note="FLAM A repeat: matches 1. .133 of consensus"
38707. .39005 /note="Alu" repeat: matches 1. .299 of consensus"
39131. .39428 /note="Alu" repeat: matches 1. .298 of consensus"
40294. .40387 /note="MIR repeat: matches 196. .103 of consensus"
```

```
repeat_region 40789. .41091 /note="Alu" repeat: matches 303. .1 of consensus"
repeat_region 41955. .42124 /note="MIR repeat: matches 178. .1 of consensus"
repeat_region 42280. .42573 /note="MIR repeat: matches 178. .1 of consensus"

Query Match 13.7%; Score 34.4; DB 9; Length 113872;
Best Local Similarity 54.8%; Pred. No. 2.7;
Matches 68; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 90 CCTGCTGACAGAGTGTGCTGGGCTGCTGGTGGGATGGATTTGACAGTCTGC 149
DB 79794 CCGGCTGCAGACAGCAGCTCTGGAGCTGGAATGGCAAGGCTGCTTTTACACAAA 79853
QY 150 CCATGTACAAAGTAACTCATATCAGATTAACAGTGTGACAGGAGACACTGCCAA 209
DB 79854 CCACACACACACACACACACACACACACACACACACACAGATACACAGACACACACA 79913
QY 210 AAGA 213
DB 79914 GACA 79917

RESULT 14
AC007273 147724 bp DNA HTG 09-APR-1999
LOCUS Homo sapiens clone NH007910. WORKING DRAFT SEQUENCE. 6 unordered
pieces.
ACCESSION AC007273
NID 94580507
VERSION AC007273.1 GI:4580507
KEYWORDS HTG; HTGS_Phasel.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutharia; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 147724)
REFERENCE Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-APR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2826: contig of 2826 bp in length
2827 2844: gap of unknown length
2845 9247: contig of 6403 bp in length
9248 9265: gap of unknown length
9266 25233: contig of 15968 bp in length
25234 41277: contig of unknown length
41278 41294: contig of 16027 bp in length
41295 89779: gap of unknown length
89780 89796: gap of unknown length
89797 147724: contig of 57928 bp in length.
location/Qualifiers
1. 147724
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH007910"
BASE COUNT 47271 a 27212 c 27349 g 45805 t 87 others
ORIGIN
```



This Page Blank (uspto)

---



KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 452)  
Hiller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevasaks, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
Insert Size: 937  
High quality sequence stops: 338  
Source: IMAGE Consortium, LBNL  
This clone is available royalty-free through LBNL; contact the  
IMAGE Consortium (info@image.lbnl.gov) for further information.  
Insert Length: 937 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 338.  
Location/Qualifiers  
1. 452  
/organism="Homo sapiens"  
/db\_xref="GDB:3771346"  
/db\_xref="taxon:9606"  
/clone="IMAGE:202215"  
/clone\_lib="Soares fetal liver spleen INF15"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
15' AACGGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'1'.  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 117 a 95 c 87 g 147 t 6 others  
ORIGIN

Query Match 6.2%; Score 18; DB 23; Length 452;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 147 AGCTGTCCTCAAAATGACAG 164  
Db .62 AGCTGTCCTCAAAATGACAG 79

RESULT 2  
LOCUS AA243037 382 bp mRNA EST 11-MAR-1998  
DEFINITION z124905.s1 StrataGene NT2 neuronal precursor 937230 Homo sapiens  
CDNA clone IMAGE:664376 3' similar to gb:U20868 60S RIBOSOMAL  
PROTEIN L4 (HUMAN);, mRNA sequence.  
ACCESSION AA243037  
NID 91873833  
KEYWORDS AA243037.1 GI:1873833  
EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS 1 (bases 1 to 382)  
Hiller, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On May 5, 1995 this sequence version replaced gi:798381.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LBNL; contact the  
IMAGE Consortium (info@image.lbnl.gov) for further information.  
Insert Length: 1485 Std Error: 0.00  
Seq primer: -41ml3 fwd. Et from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 382  
/organism="Homo sapiens"  
/db\_xref="GDB:5426364"  
/db\_xref="taxon:9606"  
/map="21"  
/clone="IMAGE:664376"  
/clone\_lib="Stratagene NT2 neuronal precursor 937230"  
/tissue\_type="neuroepithelial cells"  
/dev\_stage="Ntera-2 neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: brain; Vector: pLuscript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
oligo dt. Uninduced, exponentially growing neuroepithelial  
cells (Ntera-2/cl.D). Average insert size: 1.0 kb.  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GATTCGCGACGAC  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

BASE COUNT 72 a 102 c 90 g 118 t  
ORIGIN

Query Match 6.2%; Score 18; DB 30; Length 382;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 144 AGGAGCTGTCCAAATGA 161  
Db 35 AGGAGCTGTCCAAATGA 52

RESULT 3  
LOCUS AA771269/c 434 bp mRNA EST 29-JAN-1998  
DEFINITION vt17d09.r1 Barstead mouse myotubes MRLR5 Mus musculus CDNA clone  
IMAGE:1163345 5', mRNA sequence.  
ACCESSION AA771269  
NID 92823080  
KEYWORDS AA771269.1 GI:2823080  
EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 434)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE The WashU-HMT Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407528.



Contact: Maira M/Mouse EST Project  
 Washu-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LINT; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:629257

Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 415.

## FEATURES

## source

Location/Qualifiers

1. 434

/organism="Mus musculus"

/strain="C3H"

/db\_xref="taxon:10090"

/map="3"

/clone="IMAGE:1163345"

/clone\_lib="Barstead mouse myotubes MPLRB5"

/cell\_line="C2C12"

/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5'  
 TGTTCGAAATCGAAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 [AATTCGATCCTTG], digested with Not I and cloned into the  
 Not I and Eco RI sites of the modified pT73 vector.  
 Library constructed by Bob Barstead. The C2C12 cell line  
 (available from ATCC, catalog # CRL-1772) differentiates  
 rapidly, forming contractile myotubes and producing  
 characteristic muscle proteins."

BASE COUNT 83 a 124 c 133 g 94 t

## ORIGIN

Query Match 6.28; Score 18; DB 38; Length 434;  
 Best Local Similarity 100.0%; Pred. No. 6.3;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 CACCACTCGCTTGCGG 121  
 11111111111111111111  
 Db 160 CACACAGTCGCTTGCGG 143

## RESULT 4

## LOCUS

M78894 314 bp mRNA EST 26-MAY-1992

## DEFINITION

EST01042 Subtracted Hippocampus, Striatum (cat. #936205) Homo

## ACCESSION

## NID

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

Other ESTs: EST01041  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423

Email: arkerlav@tigr.org  
 Seq primer: M13 Reverse.

## FEATURES

## source

Location/Qualifiers

1. 314

/organism="Homo sapiens"

/db\_xref="ATCC (dna):78364D"

/db\_xref="ATCC (lnhost):78364"

/db\_xref="GDB:DIS330E"

/db\_xref="taxon:9606"

/clone="HCCPB34"

/clone\_lib="Subtracted Hippocampus, Striatum (cat.  
 #936205)"

/note="Vector: lambdaZAP-II; The hippocampus library (#4  
 above) was subtracted with a fibroblast cell line cDNA  
 library (Striatum cat. #936209; W138 lung fibroblast  
 cell line; oligo-dT + random primed cDNA synthesis;  
 lambdaZAP-II vector, 1.0kb average insert size.) by the  
 method of Sive & St. John (Nucl. Acids Res. 16:10937,  
 1988)."

BASE COUNT 72 a 79 c 57 g 106 t

## ORIGIN

Query Match 5.98; Score 17; DB 20; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 136 AGAAGTAGAGAGAGCTGT 152  
 11111111111111111111  
 Db 178 AGAAGTAGAGAGAGCTGT 194

## RESULT 5

## LOCUS

T16657 272 bp mRNA EST 25-JUL-1996

## DEFINITION

NIB1651 Normalized infant brain, Bento Soares Homo sapiens CDNA

## ACCESSION

## NID

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

Contact: Sikele UM  
 Department of Pharmacology  
 University of Colorado Health Sciences Center  
 Box C236, 4200 E. 9th Ave, Denver CO 80262-0236  
 Tel: 3032708637  
 Fax: 3032707097  
 Email: nikki@etally.uchsc.edu  
 Seq primer: -21M13 Universal.  
 Location/Qualifiers

## FEATURES

## source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Normalized infant brain, Bento Soares"  
 /lab\_host="E. coli DH5-alpha"  
 /note="Vector: BA, M13-derived; Site\_1: HindIII; Site\_2:  
 NotI; The normalized infant brain library, constructed by  
 Bento Soares, Columbia University, was oligo-(dT) primed  
 and directionally cloned into an M13-derived plasmid using  
 total brain mRNA from a 72-day old human female afflicted

with spinal muscular atrophy. The library was normalized as described elsewhere."

BASE COUNT 58 a 64 c 55 g 95 t

ORIGIN

Query Match 5.9%; Score 17; DB 20; Length 272;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 136 AGAAGTACAGAGAGCTGT 152  
|||||

Db 191 AGAAGTACAGAGAGCTGT 207

RESULT 6  
T32493 263 bp mRNA EST 06-SEP-1995  
LOCUS EST49387 Human Gall bladder Homo sapiens cDNA 3' end similar to  
DEFINITION None, mRNA sequence.  
ACCESSION T32493  
NID 9614591  
VERSION T32493.1 GI:614591  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 263)  
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulder,R.A.,  
Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,  
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S.,  
Kelley,J.M., Klinek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,  
Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,  
Pellegino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,  
Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R.,  
Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A.,  
Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,  
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,  
Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,  
Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wang,J., Xu,C.,  
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,  
Hasselthine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.

TITLE Initial Assessment of Human Gene Diversity and Expression Patterns  
JOURNAL Based Upon 83 Million Basepairs of cDNA Sequence  
MEDLINE Nature 377, 3-174 (1995)  
COMMENT 96026280

Other-ESTs: EST49388 THC13013

CONTACT: Venter, JC  
The Institute for Genomic Research  
932 Clopper Rd, Gathersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423

EMAIL: tdbinfo@tdb.tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please contact the TIGR Database  
(tdbinfo@tdb.tigr.org)  
Seq primer: M13-21.

FEATURES  
source Location/Qualifiers  
1..263

/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):103035"  
/db\_xref="taxon:9606"  
/map="21"

BASE COUNT 66 a 68 c 49 g 75 t 5 others  
ORIGIN

Query Match 5.9%; Score 17; DB 20; Length 263;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 136 AGAAGTACAGAGAGCTGT 152  
|||||

Db 151 AGAAGTACAGAGAGCTGT 167

RESULT 7  
T35391 324 bp mRNA EST 06-SEP-1995  
LOCUS EST83977 Human Parathyroid gland Homo sapiens cDNA 5' end similar  
DEFINITION to None, mRNA sequence.  
ACCESSION T35391  
NID 9617489  
VERSION T35391.1 GI:617489  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 324)  
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulder,R.A.,  
Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,  
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S.,  
Kelley,J.M., Klinek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,  
Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,  
Pellegino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,  
Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R.,  
Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A.,  
Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,  
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,  
Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,  
Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wang,J., Xu,C.,  
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,  
Hasselthine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.

TITLE Initial Assessment of Human Gene Diversity and Expression Patterns  
JOURNAL Based Upon 83 Million Basepairs of cDNA Sequence  
MEDLINE Nature 377, 3-174 (1995)  
COMMENT 96026280

Other-ESTs: THC13013

CONTACT: Venter, JC  
The Institute for Genomic Research  
932 Clopper Rd, Gathersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423

EMAIL: tdbinfo@tdb.tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please contact the TIGR Database  
(tdbinfo@tdb.tigr.org)  
Seq primer: M13 Reverse.

FEATURES  
source Location/Qualifiers  
1..324

/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):106646"  
/db\_xref="taxon:9606"  
/clone\_lib="Human Parathyroid gland"  
/note="Organ: parathyroid gland"

BASE COUNT 71 a 75 c 63 g 115 t

ORIGIN

Query Match 5.9%; Score 17; DB 20; Length 324;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 136 AGAAGTACAGAGAGCTGT 152  
|||||

Db 198 AGAAGTACAGAGAGCTGT 214

RESULT 8  
 LOCUS T52594 313 bp mRNA EST 06-FEB-1995  
 DEFINITION yb21c03.r1 Stratagene fetal spleen (#937205) Homo sapiens CDNA  
 clone IMAGE:71812 5', mRNA sequence.  
 ACCESSION T52594  
 NID 9634454  
 VERSION T52594.1 GI:654454  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 313)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Chissoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
 Maris, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevasakis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.,  
 and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 97044478  
 TITLE JOURNAL  
 MEDLINE  
 COMMENT  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 High quality sequence stops: 170  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seg primer: M13Rpi  
 High quality sequence stop: 170.  
 Location/Qualifiers  
 1. 313  
 /organism="Homo sapiens"  
 /db\_xref="GDB:493477"  
 /db\_xref="taxon:9606"  
 /map="22"  
 /clone="IMAGE:71812"  
 /clone\_1lb="Stratagene fetal spleen (#937205)"  
 /tissue\_type="fetal spleen"  
 /dev\_stage="fetal"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: spleen; Vector: pBluescript SK-; site\_1:  
 EcoRI; site\_2: XhoI; Cloned unidirectionally. Primer:  
 0190 dt. Pooled spleens. Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGCAGAG  
 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."
 BASE COUNT 45 a 86 c 77 g 103 t 2 Others  
 ORIGIN  
 Query Match 5.9%; Score 17; DB 20; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1.4.5-TRISPHOSPHATE-BINDING PROTEIN TYPE 2 RECEPTOR ;, mRNA  
 sequence.  
 R09245  
 NID 9761168  
 VERSION R09245.1 GI:761168  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 318)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P., and  
 Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)  
 TITLE JOURNAL  
 COMMENT  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 Insert Size: 783  
 High quality sequence stops: 128 Source: IMAGE Consortium, LNL This  
 clone is available royalty-free through LNL; contact the IMAGE  
 Consortium (info@image.llnl.gov) for further information.  
 Insert length: 783 Std Error: 0.00  
 Seg primer: -21m13  
 High quality sequence stop: 128.  
 Location/Qualifiers  
 1. 318  
 /organism="Homo sapiens"  
 /db\_xref="GDB:480181"  
 /db\_xref="taxon:9606"  
 /map="16"  
 /clone="IMAGE:128020"  
 /clone\_1lb="Soares fetal liver spleen INFLS"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)  
 with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
 1st strand cDNA was primed with a Pac I - oligo(dT) primer  
 15' AACGGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'}.  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified p773 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M.Fatima Bonaldo."
 BASE COUNT 97 a 77 c 67 g 77 t  
 ORIGIN  
 Query Match 5.9%; Score 17; DB 21; Length 318;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9  
 LOCUS R09245 318 bp mRNA EST 05-APR-1995  
 DEFINITION yf26g03.s1 Soares fetal liver spleen INFLS Homo sapiens CDNA clone  
 IMAGE:128020 3' similar to SP:IP3S\_RAT P29955 INOSITOL

RESULT 10  
 LOCUS D52818/c 311 bp mRNA EST 14-SEP-1995  
 DEFINITION HUM089502B Clontech human fetal brain polyA+ mRNA (#6353) Homo  
 sapiens CDNA clone GEN-089502 5', mRNA sequence.  
 ACCESSION D52818  
 NID 9953054  
 VERSION D52818.1 GI:953054

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
Takahashi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
Mekawa,H., Shin,S. and Nakamura,Y.
Fujiwara et al.(1995)
JOURNAL Unpublished (1995)
COMMENT On Nov 29, 1993 this sequence version replaced gi:503005.

TITLE
JOURNAL
COMMENT

FEATURES
source
Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawachu-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.
location/Qualifiers
1..311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GEN-089802"
/clone_lib="Clontech human fetal brain polyA+ mRNA
(#6535)"

BASE COUNT 98 a 58 c 79 g 75 t 1 others
ORIGIN

Query Match 5.9%; Score 17; DB 23; Length 311;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 136 AGAGCTAGAGAGCTGT 152
|||||
Db 139 AGACTAGAGAGCTGT 123

RESULT 11
LOCUS L38062 333 bp mRNA EST 07-NOV-1996
DEFINITION BNAF0409E Mustard flower buds Brassica rapa cDNA clone F0409, mRNA
sequence.
ACCESSION L38062
NID 9887237
VERSION L38062.1 GI:887237
KEYWORDS EST.
SOURCE Brassica rapa.
ORGANISM Brassica rapa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; eudicotyledons; core
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Brassica.
1 (bases 1 to 333)
Lim,C.O., Kim,H.Y., Kim,M.G., Lee,S.I., Chung,W.S., Park,S.H.,
Hwang,I. and Cho,M.J.
Expressed sequence tags of Chinese cabbage flower bud cDNA
Plant Physiol. 111 (2), 577-588 (1996)
96255495
On May 9, 1995 this sequence version replaced gi:803855.

TITLE
JOURNAL
COMMENT

FEATURES
source
Contact: Lim,C.O., Kim,M.G., Hwang,I. and Cho,M.J.
Plant Molecular Biology and Biotechnology Research Center
Email: pmbdr@hongae.gnu.ac.kr.
location/Qualifiers
1..333
/organism="Brassica rapa"
/strain="pekinensis"
/db_xref="taxon:3711"
/clone="F0409"
/clone_lib="Mustard flower buds"

```

```

BASE COUNT      92 a      52 c      77 g      112 t
ORIGIN

Query Match      5.9%; Score 17; DB 23; Length 333;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

Cy 184 TTCACATCAAGTTCAGA 200
|||||
Db 36 TTCAGATCAAGTTCAGA 52

RESULT 12
LOCUS      NA1057
DEFINITION yys396.g6.s1 Soares_multiple_sclerosis_2NBHMP Homo sapiens cDNA
            clone IMAGE:277306 3', mRNA sequence.
ACCESSION  NA1057
            g1164655
VERSION     NA1057.1 GI:1164655
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 365)
            Haller,L., Clark,N., Dubnue,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,D., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Tevaskis,E., Waterston,R., Williamson,A., Woldmann,P. and
            Wilson,R.
            The Mashu-Merck EST Project
            Unpublished (1995)
COMMENT     On May 5, 1995 this sequence version replaced gi:797873.

TITLE
JOURNAL

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 174
Source: IMAGE Consortium, LMLN
This clone is available royalty-free through LMLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: m13 -40 forward
High quality sequence stop: 174.
Location/Qualifiers
1..365
/organism="Homo sapiens"
/db_xref="GDB:3895682"
/db_xref="taxon:9606"
/clone_image="IMAGE:277306"
/clone_id="Soares_multiple_sclerosis_2NBHMP"
/sex="male"
/tissue_type="multiple_sclerosis_lesions"
/dev_stage="Age 46"
/alt_host="DH10B (ampicillin resistant)"
/notes="Vector: pUT73D (Pharmacia) with a modified
polylinker V_type: phagemid; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo (dt)
primer [5',
TGTTCACATCTGTAAGTGGAGGCGCCGCAATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pUT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH). "
```

BASE COUNT 75 a 92 c 62 g 129 t 7 others  
ORIGIN

Query Match 5.9%; Score 17; DB 24; Length 365;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 AGAAGTAGAGGAGCTGT 152  
|||||  
Db 208 AGAAGTAGAGGAGCTGT 224

RESULT 13  
N66420 313 bp mRNA EST 08-MAR-1996  
LOCUS  
DEFINITION Y240C05.s1 Morton Fetal Cochlea Homo sapiens cDNA clone  
IMAGE:285512 3', mRNA sequence.

ACCESSION N66420  
NID 91218545  
VERSION N66420.1 GI:1218545  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,M.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,  
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.  
and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT On May 5, 1995 this sequence version replaced gi:797648.

FEATURES  
source  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: ml3 -40 forward.

FEATURES  
source  
1. .313  
/organism="Homo sapiens"  
/db\_xref="GDB:3892629"  
/db\_xref="taxon:9606"  
/clone="IMAGE:285512"  
/clone\_1lb="Morton Fetal Cochlea"  
/tissue\_type="Cochlea"  
/dev\_stage="16-22 week fetus"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: ear; Vector: pBluescript SK-; Site:1: EcoRI;  
Site:2: XhoI; Reference: Genomics 23, 42-50 (1994) cloned  
unidirectionally. Primer: 01190 dT. Fetal cochlea, normal.  
3'x of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP  
XR vector. Library constructed by N. Robertson, C. Morton.  
-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

BASE COUNT 69 a 73 c 60 g 111 t  
ORIGIN

Query Match 5.9%; Score 17; DB 25; Length 313;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 AGAAGTAGAGGAGCTGT 152  
|||||  
Db 201 AGAAGTAGAGGAGCTGT 217

RESULT 14  
N93765 474 bp mRNA EST 20-AUG-1996  
LOCUS  
DEFINITION zb64c07.s1 Soares\_fetal\_lung\_NBH119W Homo sapiens cDNA clone  
IMAGE:308364 3', mRNA sequence.

ACCESSION N93765  
NID 91266074  
VERSION N93765.1 GI:1266074  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,M.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,  
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.  
and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT On Oct 18, 1995 this sequence version replaced gi:1026889.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 793 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 356.

FEATURES  
source  
1. .474  
/organism="Homo sapiens"  
/db\_xref="GDB:1251777"  
/db\_xref="taxon:9606"  
/clone="IMAGE:308364"  
/clone\_1lb="Soares\_fetal\_lung\_NBH119W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: pTZ19 (Pharmacia) with a  
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTCCATCTGAGTGGGAGCGGCGCCCATTTTTTTTTTTT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pTZ19 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Benito  
Soares and M. Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NBH119W."

BASE COUNT 103 a 127 c 78 g 164 t 2 others  
ORIGIN

Query Match 5.9%; Score 17; DB 25; Length 474;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 136 AGAAGTAGAGGAGCTGT 152  
|||||

Db 211 AGACTAGAGAGCTGT 227

RESULT 15

LOCUS M44889 377 bp mRNA EST 10-OCT-1996  
DEFINITION zc05908.r1 Soares.parathyroid\_tumor\_NDHPA Homo sapiens cDNA clone  
IMAGE:321470.5 similar to gb:M28212 RAS-RELATED PROTEIN RAB-6  
(HUMAN);, mRNA sequence.

ACCESSION

M44889

NID

91328989

VERSION

M44889.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 377)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Riklin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

TITLE

The Wasnu-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

On Sep 21, 1992 this sequence version replaced gi:279339.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 2660 Std Error: 0.00

Seq primer: mob.REGA+ET.

FEATURES

Location/Qualifiers

1..377

/organism="Homo sapiens"

/db\_xref="GDB:1259128"

/db\_xref="taxon:9606"

/clone="IMAGE:321470"

/clone\_lib="Soares.parathyroid\_tumor\_NDHPA"

/tissue\_type="parathyroid tumor"

/dev\_stage="adult"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: parathyroid gland; Vector: pT7T3D

(Pharmacia) with a modified polylinker; Site 1: Not I;

Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer

15-

TGTTACCAATCTGAAGTGGAGCGCGCCGACCAATTTTTTTTTTTTTTTTTTTT

T-3', double-stranded cDNA was size selected, ligated to

Eco RI adapters (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of a modified pT7T3

vector (Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M.Fatima Bonaudo. RNA from sporadic parathyroid

adenomas was kindly provided by Dr. Stephen Marx, National

Institute of Diabetes and Digestive and Kidney Diseases,

NIH"

BASE COUNT 92 a 109 c 84 g 89 t 3 others  
ORIGIN

Query Match 5.9%; Score 17; DB 26; Length 377;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 GAACCTAGAGAGAGCTGTC 153

DB 257 GAACCTAGAGAGAGCTGTC 273

Search completed: August 6, 1999, 09:22:40  
Job time: 6231 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:33 ; Search time 3059.95 Seconds  
(without alignments)  
162.447 Million cell updates/sec

Title: US-09-049-696-6  
Perfect score: 252  
Sequence: 1 CAAGAAATGTGTGTAGT.....GAGGACGCTCCATCTGCAGC 252

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	50	19.8	480	35	AA581198	AA581198 nd38c07.r
2	20	7.9	574	40	AA944027	AA944027 EST199526
3	20	7.9	673	40	AA944028	AA944028 EST199527
4	18	7.1	327	23	F15082	F15082 SSC6A10.POT
5	17	6.7	289	20	Z44307	Z44307 HSC1X021.n
6	17	6.7	482	23	H57955	H57955 Y12H03.s1
7	17	6.7	393	23	L11850	L11850 MUSEST34A.M
8	17	6.7	299	25	W00254	W00254 T1110.MVAT4
9	17	6.7	426	28	AA089192	AA089192 mo05g12.r
10	17	6.7	347	28	AA089208	AA089208 mo05g12.r
11	17	6.7	391	28	AA111421	AA111421 mo54e01.r
12	17	6.7	405	29	AA138647	AA138647 mr62g10.r
13	17	6.7	437	29	AA154648	AA154648 mt65b11.r
14	17	6.7	360	29	AA174336	AA174336 mo96g10.r
15	17	6.7	497	30	AA200290	AA200290 mu04c10.r
16	17	6.7	557	35	AA538011	AA538011 vj34h08.r
17	17	6.7	402	36	AA636730	AA636730 vr20d05.r
18	17	6.7	436	36	AA637862	AA637862 vr29f03.r
19	17	6.7	422	37	AJ003273	AJ003273 AJ003273
20	17	6.7	505	38	AA756082	AA756082 vv37e04.r
21	17	6.7	449	38	AA794783	AA794783 vu64a04.r
22	17	6.7	492	38	AA796934	AA796934 vs96a08.r
23	17	6.7	332	40	AA919265	AA919265 vs51a09.r
24	17	6.7	415	40	AA930427	AA930427 vs58g03.r
25	17	6.7	607	43	AI172274	AI172274 EST218275
26	17	6.7	446	43	AI193123	AI193123 qe69h11.x
27	17	6.7	520	45	AA818634	AA818634 UI-R-A0-a
28	17	6.7	536	46	AI413205	AI413205 mc26a08.x
29	17	6.7	286	46	AI429956	AI429956 vv37e04.y
30	17	6.7	483	48	AI550845	AI550845 mo05c12.y
31	17	6.7	434	48	AI573362	AI573362 ud42c09.y
32	17	6.7	553	48	AI601470	AI601470 fco7b09.x
33	17	6.7	441	48	AI605409	AI605409 mr62g10.y
34	16	6.3	183	20	T119166	T119166 e05004s.Tes
35	16	6.3	270	21	R13675	R13675 YF64a05.r1
36	16	6.3	511	21	T67851	T67851 YC27g08.r1
37	16	6.3	382	21	T79526	T79526 Yd45d11.r1
38	16	6.3	429	21	T81582	T81582 Yd44b02..s1
39	16	6.3	462	22	H02429	H02429 YJ40f08..r1
40	16	6.3	386	22	H13304	H13304 YJ72a03.r1
41	16	6.3	431	22	H13359	H13359 YJ71a09.r1
42	16	6.3	471	22	R31469	R31469 YH75a12..r1
43	16	6.3	415	22	R43074	R43074 YG17g08..s1
44	16	6.3	466	22	R44282	R44282 YG35b07..s1
45	16	6.3	457	50	AI668580	AI668580 YJ14g11.x

ALIGNMENTS

RESULT 1  
AA581198 480 bp mRNA EST 05-JAN-1998  
LOCUS nd38c07.r1 NCI\_CGAP\_Col Homo sapiens cDNA clone IMAGE:802572 5'  
DEFINITION similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.  
ACCESSION AA581198  
NID 92358970

```

VERSION      AA581198.1  GI:2358970
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 480)
              NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
              Unpublished (1997)
JOURNAL      On May 9, 1995 this sequence version replaced gi:803001.
COMMENT      Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert.Strausberg@nih.gov
              Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael
              Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: Life Technologies Inc., David Krizman,
              Ph.D.
              cDNA Library Arraying: Greg Lennon, Ph.D.
              DNA sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              www-bio.11nl.gov/dbip/image/image.html

Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 405.
Location/Qualifiers
1..480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:802572"
/clone_id="NCI.CGAP.C01"
/tissue_type="bulk tumor"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT2; Site_1: Salt;
       Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt."

BASE COUNT   168 a 101 c 97 g 114 t
ORIGIN

Query Match      19.88; Score 50; DB 35; Length 480;
Best Local Similarity 100.0%; Pred. No. 3; Gc-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGAATTGCTGTTAGCTTCACAAATCTGAGACATGGCGACTGG 50
    |||||||
Db 391 CAAGAATTGCTGTTAGCTTCACAAATCTGAGACATGGCGACTGG 440

RESULT 2
LOCUS      AA944027      574 bp      mRNA      EST      16-JUN-1998
DEFINITION EST199526 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
REMARKS    AA944027 end, mRNA sequence.
ACCESSION  AA944027
NID        93103943
VERSION    AA944027.1  GI:3103943
KEYWORDS   EST.
SOURCE     Rattus sp.
ORGANISM   Rattus sp.
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
              Lee, N.H., Glodex, A., Chandra, I., Mason, T.M., Quackenbush, J.,
              Kerlavage, A.R. and Adams, M.D.
              Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
              Gene Index
              Unpublished (1998)
JOURNAL      On Jan 14, 1998 this sequence version replaced gi:1796868.
COMMENT      Other_ESTs: EST199527
              Contact: Lee, NH

```

```

ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..574
/organism="Rattus sp."
/db_xref="ATCC (inhost):2013516"
/db_xref="taxon:10118"
/map="6; 22"
/clone_image="REMAA44"
/clone_id="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT   140 a 130 c 159 g 145 t
ORIGIN

Query Match      7.9%; Score 20; DB 40; Length 574;
Best Local Similarity 100.0%; Pred. No. 0.88;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 TGGATGCTGACATTGACA 142
    |||||||
Db 549 TGGATGCTGACATTGACA 568

RESULT 3
LOCUS      AA944028/c      673 bp      mRNA      EST      16-JUN-1998
DEFINITION EST199527 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
REMARKS    AA944028 5' end similar to actin bundling protein, mRNA sequence.
ACCESSION  AA944028
NID        93103944
VERSION    AA944028.1  GI:3103944
KEYWORDS   EST.
SOURCE     Rattus sp.
ORGANISM   Rattus sp.
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
              Lee, N.H., Glodex, A., Chandra, I., Mason, T.M., Quackenbush, J.,
              Kerlavage, A.R. and Adams, M.D.
              Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
              Gene Index
              Unpublished (1998)
JOURNAL      On Jan 14, 1998 this sequence version replaced gi:1796869.
COMMENT      Other_ESTs: EST199526
              Contact: Lee, NH
              ATCC
              The Institute for Genomic Research
              9712, Medical Center Drive, Rockville, MD 20850, USA
              Tel: (301)-838-3529
              Fax: (301)-838-0208
              Email: nhlee@tigr.org
              Seq primer: M13 Reverse.
              Location/Qualifiers
1..673
/organism="Rattus sp."
/db_xref="ATCC (inhost):2013516"
/db_xref="taxon:10118"
/map="14"
/clone_image="REMAA44"
/clone_id="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT   145 a 197 c 169 g 162 t
ORIGIN

```



Query Match 7.9%; Score 20; DB 40; Length 673;  
 Best Local Similarity 100.0%; Pred. No. 0.89;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 TGGGATGGTGCATTGGACA 142  
 ||||||||||||||||  
 Db 543 TGGGATGGTGCATTGGACA 524

RESULT 4  
 F15082 327 bp mRNA EST 09-SEP-1996  
 LOCUS SSC6A10 Porcine small intestine cDNA library Sus scrofa cDNA clone  
 DEFINITION cga10, mRNA sequence.  
 ACCESSION F15082  
 NID 9972063  
 VERSION F15082.1 GI:972063  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Cetartiodactyla; Suidae; Sus.  
 1 (bases 1 to 327)  
 Wintero,A.K., Fredholm,M. and Davies,W.  
 Evaluation and characterization of a porcine small intestine cDNA  
 library: analysis of 839 clones  
 Mamm. Genome 7 (7), 509-517 (1996)  
 96327607  
 COMMENT On May 8, 1995 this sequence version replaced gi:799705.

FEATURES  
 source  
 1..327  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone="cga10"  
 /clone\_lib="Porcine small intestine cDNA library"  
 /note="directionally cloned cDNA in X1.1-blue MRF"

BASE COUNT 78 a 82 c 91 g 72 t 4 others

ORIGIN  
 Contact: A.K. Wintero  
 Department of Animal Science and Animal Health, Division of Animal  
 Genetics, The Royal Veterinary and Agricultural University  
 Bulowsvej 13, 1870 Frederiksberg C, Denmark.  
 Location/Qualifiers

Query Match 7.1%; Score 18; DB 23; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 CGACTGATCAGCAGGC 81  
 ||||||||||||||||  
 Db 61 CGACTGATCAGCAGGC 78

RESULT 5  
 Z44307 289 bp mRNA EST 14-NOV-1994  
 LOCUS HSC1X021 normalized infant brain cDNA Homo sapiens cDNA clone  
 DEFINITION c-1x02, mRNA sequence.  
 ACCESSION Z44307  
 NID 9573433  
 VERSION Z44307.1 GI:573433  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 289)  
 Auffray,C., Benar,G., Bois,F., Bouchier,C., da Silva,C.,  
 Devignes,M.D., Duprat,S., Houlgate,R., Jumeau,M.N., Lamy,B.,  
 Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,  
 Sebastienl-Kabakchis,C. and Tessier,A.  
 IMAGE: molecular integration of the analysis of the human genome

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

and its expression  
 C.R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 95277534

JOURNAL  
 MEDLINE  
 COMMENT

Contact: Genethon  
 Genexpress-genethon  
 Genethon Centre de recherche sur le Genome Humain  
 1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
 Tel: 33169472800  
 Fax: 33160778698  
 Email: genexpress@genethon.fr  
 Single read  
 Genexpress\_library\_id: C; Genexpress\_sequence\_id: ylc-1xe02  
 Seq primer: (-21)M3-Universal.  
 Location/Qualifiers

FEATURES  
 source  
 1..289  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="c-1xe02"  
 /clone\_lib="normalized infant brain cDNA"  
 /sex="Female"  
 /tissue\_type="total brain"  
 /dev\_stage="3 months old"  
 /note="Organ: brain; Vector: lambda B; Site 1: HindIII;  
 Site 2: NotI; sex: Female; dev\_stage=3 months old;  
 isolate=muscular atrophy patient; tissue\_type=total  
 brain; total mRNA was oligo-(dT) primed and directionally  
 cloned 5' -> 3' into the HindIII -> NotI sites of the  
 lambda B vector. Clone library from B.Souares, Psychiatry  
 Dept. Columbia University, USA. Normalization method:  
 Bento Soares, P.N.A.S. in press"

BASE COUNT 57 a 80 c 83 g 65 t 4 others

ORIGIN

Query Match 6.7%; Score 17; DB 20; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 GCACCTTCAGGAGGAC 239  
 ||||||||||||||||  
 Db 141 GCACCTTCAGGAGGAC 157

RESULT 6  
 H57955 482 bp mRNA EST 05-OCT-1995  
 LOCUS YR12H03.s1 Soares fetal liver spleen INFIS Homo sapiens cDNA clone  
 DEFINITION IMAGE:205109 3 similar to contains A3R repetitive element ;, mRNA  
 sequence.  
 ACCESSION H57955  
 NID 91010787  
 VERSION H57955.1 GI:1010787  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 482)  
 Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
 Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
 Wilson,R.  
 The WashU-Merck EST Project  
 Unpublished (1995)  
 On Jan 25, 1995 this sequence version replaced gi:637645.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: estewartson.wustl.edu  
 Insert Size: 1003  
 High quality sequence stops: 370  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 1003 Std Error: 0.00  
 Seq primer: Promega -21m13  
 High quality sequence stop: 370.

## FEATURES

source

1..482  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3774240"  
 /db\_xref="taxon:9606"  
 /map="6"  
 /clone="IMAGE:205109"  
 /clone.lib="Soares fetal liver spleen INFs"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
 with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
 1st strand cDNA was primed with a Pac I - oligo(dT) primer  
 [5' AACGGAGATTAATTAAGATCTTTTTTTTTTTTTTTT 3']  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M. Fatima Bonaldo."

## BASE COUNT

119 a 105 c 124 g 131 t 3 others

## ORIGIN

Query Match 6.7%; Score 17; DB 23; Length 482;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 GCAGCTTCAGGAGGAC 239

Db 143 GCAGCTTCAGGAGGAC 127

## RESULT 7

L11850/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

11850 393 bp mRNA EST 21-AUG-1995  
 MUSEST34A Mouse liver Mus musculus domesticus cDNA clone ml34, mRNA  
 sequence.  
 11850 1193117  
 11850.1 GI:193117  
 EST.  
 western European house mouse.  
 Mus musculus domesticus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 393)  
 Warden,C.H., Mehraabian,M., He,K.Y., Yoon,M.Y., Diep,A., Xia,Y.R.,  
 Wen,P.Z., Svenson,K.L., Sparks,R.S. and Luisi,A.J.  
 Linkage mapping of 40 randomly isolated liver cDNA clones in the  
 mouse  
 Genomics 18, 295-307 (1993)  
 94116997  
 On May 8, 1995 this sequence version replaced gi:801268.

Contact: Craig H. Warden  
 Department of Medicine, Division of Cardiology  
 UCLA  
 Los Angeles, CA 90024-1679, USA

Locus (D5Ucl4)  
 Insert Length: 1136 Std Error: 0.00  
 High quality sequence stop: 293.  
 Location/Qualifiers  
 1..393

## FEATURES

source

/organism="Mus musculus domesticus"  
 /db\_xref="ATCC (lnhost):63272"  
 /db\_xref="taxon:10092"  
 /clone="ml34"  
 /clone.lib="Mouse liver"  
 /note="Mouse liver library"  
 BASE COUNT 79 a 104 c 116 g 94 t

## ORIGIN

Query Match 6.7%; Score 17; DB 23; Length 393;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 TTGAGCTGGGCTCTGG 120

Db 177 TTGAGCTGGGCTCTGG 161

## RESULT 8

W00254

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1110 MWAT4 bloodstream form of serodeme WPARat1.1 Trypanosoma  
 brucei rhodesiense cDNA 5', mRNA sequence.  
 W00254  
 91271707  
 W00254.1 GI:1271707  
 EST.  
 Trypanosoma brucei rhodesiense.  
 Trypanosoma brucei rhodesiense  
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.  
 1 (bases 1 to 299)  
 Djikeng,A., Donelson,J.E. and Majiwa,P.A.O.  
 Generation of expressed sequence tags as physical landmarks in the  
 genome of Trypanosoma brucei  
 Unpublished (1996)  
 On Apr 14, 1993 this sequence version replaced gi:693436.

Contact: Majiwa PAO  
 Molecular Biology Unit  
 International Livestock Research Institute  
 P.O. Box 30709, Nairobi, Kenya  
 Tel: 254-2 630743  
 Fax: 254-2 631499  
 Email: p.majiwa@cgiar.com  
 Seq primer: T3 primer.  
 Location/Qualifiers  
 1..299  
 /organism="Trypanosoma brucei rhodesiense"  
 /db\_xref="taxon:31286"  
 /map="9"  
 /clone.lib="MWAT4 bloodstream form of serodeme WPARat1.1"  
 /note="Vector: Lambda Zap II (Stratagene); Site\_1: EcoRI;  
 Site\_2: XhoI; The mRNA was purified from a cloned  
 population of bloodstream trypanosomes reexpressing the  
 MWAT4 metacyclic variant surface glycoprotein (VSG). A  
 unidirectional oligo dT-primed EcoRI/XhoI cDNA library was  
 constructed in lambda Zap II (Stratagene)."

## FEATURES

source

BASE COUNT 72 a 62 c 69 g 95 t 1 others

## ORIGIN

Query Match 6.7%; Score 17; DB 25; Length 299;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 GATGGTACATTGACA 142

Db 207 GATGGTACATTGACA 223

## RESULT 9

AA089192/c  
LOCUS AA089192 426 bp mRNA EST 15-FEB-1997  
DEFINITION mo05912.r1 Strataene mouse lung 937302 Mus musculus cDNA clone  
IMAGE:552742 5', mRNA sequence.  
ACCESSION AA089192  
NID 91634687  
VERSION AA089192.1 GI:1634687  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:  
Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
1 (bases 1 to 426)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMT Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393064.  
FEATURES  
source  
Contact: Marra M/Mouse EST Project  
WashU-HMT Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:333334  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 403.  
Location/Qualifiers  
1..426  
/organism="Mus musculus"  
/strain="C57BL/6 x CBA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:552742"  
/clone\_lib="Strataene mouse lung 937302"  
/sex="female"  
/tissue\_type="lung"  
/dev\_stage="6-8 month old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site:1:  
ECORI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. 6-8 month old female lung and 1.5 year old male  
lung were source of mRNA. Average insert size: 1.5 kb;  
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGCAG  
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:  
Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
1 (bases 1 to 347)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMT Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393058.  
FEATURES  
source  
Contact: Marra M/Mouse EST Project  
WashU-HMT Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:333486  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 312.  
Location/Qualifiers  
1..347  
/organism="Mus musculus"  
/strain="C57BL/6 x CBA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:552694"  
/clone\_lib="Strataene mouse lung 937302"  
/sex="female"  
/tissue\_type="lung"  
/dev\_stage="6-8 month old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site:1:  
ECORI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. 6-8 month old female lung and 1.5 year old male  
lung were source of mRNA. Average insert size: 1.5 kb;  
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGCAG  
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

Query Match 6.7%; Score 17; DB 28; Length 347;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 104 TTGAGCTGGGCTCTGG 120  
DB 187 TTGAGCTGGGCTCTGG 171

RESULT 11  
LOCUS AA111421/c 391 bp mRNA EST 06-NOV-1996  
DEFINITION mo54e01.r1 Life Tech mouse embryo 10 5dp 10665016 Mus musculus  
CDNA clone IMAGE:557400 5', mRNA sequence.  
ACCESSION AA111421  
NID 91663302  
VERSION AA111421.1 GI:1663302  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:  
Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
1 (bases 1 to 391)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Nov 29, 1993 this sequence version replaced g1:636299.

Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG:338192  
Seq primer: -28M13 rev1 from Amersham  
High quality sequence stop: 340.

## FEATURES

SOURCE

Location/Qualifiers  
1..391  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:557400"  
/clone\_lib="Life Tech mouse embryo 10 5dpc 10665016"  
/tissue\_type="embryo"  
/dev\_stage="10.5dpc embryos"  
/lab\_host="DH10B"  
/note="Organ: whole embryo; Vector: PCW-SPORT2; Site\_1:  
-SalI; Site\_2: NotI; Cloned unidirectionally. Primer:  
Oligo dT. 10.5dpc embryos. PCW-SPORT2 vector."  
BASE COUNT 78 a 101 c 122 g 90 t  
ORIGIN

Query Match 6.7%; Score 17; DB 28; Length 391;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 TTGACCTGGGCTCTGG 120  
|||||  
Db 198 TTGACCTGGGCTCTGG 182

RESULT 12  
AA138647 405 bp mRNA EST 11-FEB-1997  
LOCUS AA138647/c  
DEFINITION m62g10.r1 Stratagene mouse testis (#937308) Mus musculus cDNA  
sequence. IMAGE:602082 5' similar to TR:G436770 G436770 FASCIN. ;, mRNA  
AA138647  
NID g1700850  
VERSION AA138647.1 GI:1700850  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 405)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced g1:1394931.

Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG:367514  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 364.

## FEATURES

SOURCE

Location/Qualifiers  
1..405  
/organism="Mus musculus"  
/strain="Inbred CD-1"  
/db\_xref="taxon:10090"  
/clone="IMAGE:602082"  
/clone\_lib="Stratagene mouse testis (#937308)"  
/sex="Males"  
/tissue\_type="testis"  
/dev\_stage="10-12 week old"  
/lab\_host="SOLR (Kanamycin resistant)"  
/note="Organ: testis; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;  
-5' adaptor sequence: 5' GATTCGGCAGCAG 3' -3' adaptor  
sequence: 5' CTCGAGTCTTTTCTTTT 3'."  
BASE COUNT 84 a 115 c 105 g 101 t  
ORIGIN

Query Match 6.7%; Score 17; DB 29; Length 405;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 TGGGATGGTGACATTG 139  
|||||  
Db 272 TGGGATGGTGACATTG 256

RESULT 13  
AA154648 437 bp mRNA EST 11-DEC-1996  
LOCUS AA154648/c  
DEFINITION m65B11.r1 Soares mouse lymph node NbM1N Mus musculus cDNA clone  
IMAGE:634749 5', mRNA sequence.  
AA154648  
NID g1726286  
VERSION AA154648.1 GI:1726286  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 437)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Apr 14, 1993 this sequence version replaced g1:693201.

Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG:386741  
putative full length read  
vector to vector length is 572  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 425.

## Location/Qualifiers

Query Match	6.7%;	Score 17;	DB 29;	Length 437;
Best Local Similarity	100.0%;	Pred. NO. 38;		
Matches 17; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

RESULT	14
AA174336/c	
LOCUS	
DEFINITION	
AA174336	360 bp mRNA
0966g10.1	Stratagene mouse testis (#937308) Mus musculus cDNA
clone IMAGE:567618 5'	mRNA sequence.

VERSION	AA174336.1	GI:1755523
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE  
AUTHORS

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia.  
Eutheria: Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 360)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuc

-TITLE The WashU-HHMI Mouse EST Project  
 JOURNAL unpublished (1996)  
 COMMENT On Apr 14, 1993 this sequence version replaced g1:692771.

Contact: Maira M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LIND; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

```

FEATURES
source      Location/Qualifiers
1. 360      /organism="Mus musculus"

```

```
Query Match      6.7%; Score 17; DB 29; Length 360;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

QY 220 GCAGCAGCTTCAGGAGG 236  
|||||  
Db 68 GCAGCAGCTTCAGGAGG 52

RESULT	15
AA000290/c	
LOCUS	
DEFINITION	
AA000290	497 bp mRNA
mu04c10.r1 Soares mouse	3NDMS Mus musculus
5', mRNA sequence.	EST
	17-FEB-1997
	clone IMAGE:658418

ACCESSION	AA200290
NID	g1796556
VERSION	AA200290.1
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE  
1 (bases 1 to 497)  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.

TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On May 8, 1995 this sequence version replaced gi:801112

**CONTACT:** Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.jnl.gov) for further information.  
 MGI:390410  
 Seq primer: -28M13 rev2 from Amerham  
 High quality sequence stop: 494.  
 Location/Qualifiers  
 1..497

```
FEATURES
source
Location/Qualifiers
1. 497
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:538418"
/clone_1ib="Soares mouse 3NDMS
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
```

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCTGAGTGGAGCGCGCGCTGTGTGTGTGTGTGTGTGT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. RNA  
provided by Dr. Bertrand Jordan. Library went through  
three rounds of normalization, and was constructed by  
Bento Soares and M.Fatima Bonaldo."

BASE COUNT 164 a 109 c 96 g 128 t  
ORIGIN

Query Match 6.7%; Score 17; DB 30; Length 497;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 134 CATTGACAGTGTCTGCC 150  
|||||  
DB 45 CATTGACAGTGTCTGCC 29

Search completed: August 6, 1999, 09:22:37  
Job time: 6228 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:10:11 ; Search time 3264.21 Seconds

(without alignments)  
233.831 Million cell updates/sec

Title: US-09-049-696-3

Perfect score: 240  
Sequence: 1 AAAATGCTGATGTTCTGCTT.....GAATTCCTATTCATG 240

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBank: \*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_v1:\*  
18: em\_fun:\*  
19: em\_hcg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl1:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_v1:\*  
34: gb\_htg1:\*  
35: gb\_htg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	240	100.0	2826	42	AF127036	Homo sapi

2	238.4	63.8	35278	11	AF039400	AF039400 Homo sapi
3	153	62.0	2937	12	AB017156	AB017156 Mus muscu
4	148.8	62.0	2937	11	AB017156	AB017156 Mus muscu
5	99.6	41.5	3415	42	AF043976	AF043976 Homo sapi
6	98	40.8	2984	3	BTU36445	BTU36445 Bos taurus
7	94.8	39.5	3317	3	AF001261	AF001261 Bos tauru
8	94.8	39.5	3288	3	AF001262	AF001262 Bos tauru
9	94.8	39.5	2820	3	AF001263	AF001263 Bos tauru
10	94.8	39.5	1265	3	AF001264	AF001264 Bos tauru
11	94.6	39.4	3022	12	AF047938	AF047938 Mus muscu
12	94.6	39.4	3471	12	AF052746	AF052746 Mus muscu
13	76	31.7	3604	9	AB026833	AB026833 Homo sapi
14	36	15.0	19436	2	U67522	U67522 Methanococ
15	1251	14.4	1251	2	MEP6K1	X64296 M.eugenii m
16	34.6	14.4	2207	17	RE03M2C	M19408 Reovirus se
17	34.6	14.4	2203	17	RE03OCPMUA	M20161 Reovirus ty
18	34.6	14.4	2203	17	RSU24260	U24260 Reovirus ty
19	34	14.2	5131	1	RIRSPAP	M37647 Rickettsia
20	34	14.2	5319	5	AR019483	AR019483 Sequence
21	33	13.8	153098	36	PRMAL3P2	AL034558 Plasmodiu
22	32.4	13.5	3054	9	GCRBGL0ETA	M54984 G.craissicu
23	32.4	13.5	57113	11	OCU60902	U60902 Otollemur cr
24	32.4	13.5	1452	42	AF131758	AF131758 Homo sapi
25	32.2	13.4	1744	10	HSPROTX	Y13145 Homo sapien
26	32.2	13.4	1949	10	HSU79296	U79296 Human dihyd
27	32.2	13.4	2320	11	AF001437	AF001437 Homo sapi
28	32.2	13.4	1581	11	AF052166	AF052166 Homo sapi
29	32.2	13.4	2365	11	HSU82328	U82328 Homo sapien
30	31.8	13.2	36262	36	CELR07EA	X93652 Caenorhabdi
31	31.6	13.2	9541	3	HSINT2HB	X86012 Human DNA s
32	31.4	13.1	1465	10	WALPGRX	J03076 Macropus ro
33	31.4	13.1	5763	4	XXU53471	U53471 Xiphophorus
34	31.4	13.1	1467	7	GLU01819	U01819 Gelidium la
35	31.4	13.1	198582	11	AC005291	AC005291 Homo sapi
36	31.4	13.1	1971	12	AF056034	AF056034 Rattus no
37	31.4	13.1	1821	12	AF056035	AF056035 Rattus no
38	31.2	13.0	1453	4	CHKRGR	L37101 Gallus gall
39	31.2	13.0	446	14	G37442	G37442 SHGC-57849
40	31.2	13.0	200301	35	AC007003	AC007003 Homo sapi
41	31	12.9	112585	9	HS105957	AL023095 Human DNA
42	31	12.9	183461	11	AC006121	AC006121 Homo sapi
43	31	12.9	179229	34	HS467K16	AL031283 Homo sapi
44	30.8	12.8	218074	11	HUAC002044	AC002044 Human chr
45	30.6	12.8	4644	1	MTDNAGRP	X58406 M.tuberculo

# ALIGNMENTS

RESULT 1	AF127036	2826 bp	MRNA	PRI	16-APR-1999
LOCUS	AF127036				
DEFINITION	Homo sapiens calcum-activated chloride channel protein 1 (CACCL1)				
ACCESSION	AF127036				
VERSION	AF127036.1				
KEYWORDS	GI:4585468				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 2826)				
JOURNAL	Agnel, M., Vernal, T. and Culouscou, J.-M.				
REFERENCE	Cloning of three human homologs of bovine epithelial chloride				
CHANNEL	channel				
UNPUBLISHED	Unpublished				
2 (bases 1 to 2826)					
REFERENCE	Agnel, M. and Culouscou, J.-M.				
AUTHORS	Direct Submission				
TITLE	Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des				
JOURNAL	Carrières, Rueil-Malmaison 92500, France				
FEATURES	Location/Qualifiers				

source 1. 2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. 2826  
/gene="CACCI1"  
5. 2749  
/gene="CACCI1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPKRSVFLLILHLLEGALNSLIQLNNNGEGIVVAIDPNVP  
EDELIIQIKDMVTOASLYLEFATGRKRYFNKVALIIPETWKADYVRPLEYKNA  
DVLVAESTPGNDPEYTEMONGCEKGERHILTPDFIGKLAEGPQGRAPVEMAH  
LRMGVPEYNNDEKRYLSNGRLOARCSAGITGNVYKCGGSCYTRCFNKVTGL  
YKCEPEYLOSROTERKASIMFRAQVDSIVERCTEQNNKKEAPNKNQNCINRSTWEI  
RDSDFKRTTPTTPPPNPTSLQIQIRIVCLVLDKSGMATGNRLNRLNQAQQLFL  
LQTVELGSMVGMVTFDSAAHVOSELIQINSGRDRLAKRLPAAASGGSISCSGLRSA  
FTVIRKRYPTDGESEIVLTDGEDNTISCFNEVKOSGAIHTVALGSAOQLELSK  
MTGGLQTYASPOVONNGLIDAFGLSSGNAGVSORSIOLESKGLTLQNSOMNGTVIY  
DSYKQDILFLITWTQPOIILMDPSGOGKGGFVVDKNTKMAVLIQIGIAKVGWKY  
SLQASSTQTLITVTSRASNAITLPIITVTSKNTKTSKPSPLVYANIRGASPIILRA  
SVTALIESVNGKIVTLELNDGAGADATKDDGVISRYFTTDTNGRISVKRALGVN  
AARRVVIPOOGSALYIPGMWENDEINONPPREINKDVOHKOVCFSRTSSGGSFVAS  
DVPNAPIPDLPPGQITDLKAEIHGSSLINTWTAPGDYDGHGAHKYIIRISTSIDL  
LRDFNENSLQVNTALIPKEANSEVEFLKPEINTFFENGTDLEIFAIQAVDVKSEI  
SNIRVSLFIPQPTPEPSPDETSAQCPNHINSTIPGHIHLKIMMKWIGELQLSIA  
"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0% Score 240; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 1.1e-63;  
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAATGCTGATGTTCTGCTGAGTCTACTCTCCACAGTATGATGAACCTACACTG 60  
Db 303 AAAATGCTGATGTTCTGCTGAGTCTACTCTCCACAGTATGATGAACCTACACTG 362

Qy 61 AGCAGATGGGCAACTGCTGAGAGAGAAAGGTTGAAGATCCACTCTCTGATTTCAATG 120  
Db 363 AGCAGATGGGCAACTGCTGAGAGAGAAAGGTTGAAGATCCACTCTCTGATTTCAATG 422

Qy 121 CAGGAAAAAGTTAGCTGAATATGACCAAGGTAGGCGATTGTGCATGATGGGGCTC 180  
Db 423 CAGGAAAAAGTTAGCTGAATATGACCAAGGTAGGCGATTGTGCATGATGGGGCTC 482

Qy 181 ATCTACGATGGGAGTATTGACGAGTACAAATATGATGAGAAATTTCTACTATCCAAATG 240  
Db 483 ATCTACGATGGGAGTATTGACGAGTACAAATATGATGAGAAATTTCTACTATCCAAATG 542

RESULT 2  
AF039400 3311 bp mRNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,  
DEFINITION complete cds.  
ACCESSION AF039400  
NID 94009457  
AF039400.1 GI:4009457  
KEYWORDS human.  
SOURCE human.  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional

analysis of human CLCA1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
Genomics 54 (2), 200-214 (1998)  
JOURNAL MEDLINE 99047526  
REFERENCE 2 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source location/Qualifiers  
1. 3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. 3311  
/gene="hCLCA1"  
352. 3096  
/gene="hCLCA1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPKRSVFLLILHLLEGALNSLIQLNNNGEGIVVAIDPNVP  
EDELIIQIKDMVTOASLYLEFATGRKRYFNKVALIIPETWKADYVRPLEYKNA  
DVLVAESTPGNDPEYTEMONGCEKGERHILTPDFIGKLAEGPQGRAPVEMAH  
LRMGVPEYNNDEKRYLSNGRLOARCSAGITGNVYKCGGSCYTRCFNKVTGL  
YKCEPEYLOSROTERKASIMFRAQVDSIVERCTEQNNKKEAPNKNQNCINRSTWEI  
RDSDFKRTTPTTPPPNPTSLQIQIRIVCLVLDKSGMATGNRLNRLNQAQQLFL  
LQTVELGSMVGMVTFDSAAHVOSELIQINSGRDRLAKRLPAAASGGSISCSGLRSA  
FTVIRKRYPTDGESEIVLTDGEDNTISCFNEVKOSGAIHTVALGSAOQLELSK  
MTGGLQTYASPOVONNGLIDAFGLSSGNAGVSORSIOLESKGLTLQNSOMNGTVIY  
DSYKQDILFLITWTQPOIILMDPSGOGKGGFVVDKNTKMAVLIQIGIAKVGWKY  
SLQASSTQTLITVTSRASNAITLPIITVTSKNTKTSKPSPLVYANIRGASPIILRA  
SVTALIESVNGKIVTLELNDGAGADATKDDGVISRYFTTDTNGRISVKRALGVN  
AARRVVIPOOGSALYIPGMWENDEINONPPREINKDVOHKOVCFSRTSSGGSFVAS  
DVPNAPIPDLPPGQITDLKAEIHGSSLINTWTAPGDYDGHGAHKYIIRISTSIDL  
LRDFNENSLQVNTALIPKEANSEVEFLKPEINTFFENGTDLEIFAIQAVDVKSEI  
SNIRVSLFIPQPTPEPSPDETSAQCPNHINSTIPGHIHLKIMMKWIGELQLSIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 99.3% Score 238.4; DB 11; Length 3311;  
Best Local Similarity 99.6%; Pred. No. 3.3e-63;  
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAATGCTGATGTTCTGCTGAGTCTACTCTCCACAGTATGATGAACCTACACTG 60  
Db 650 AAAATGCTGATGTTCTGCTGAGTCTACTCTCCACAGTATGATGAACCTACACTG 709

Qy 61 AGCAGATGGGCAACTGCTGAGAGAGAAAGGTTGAAGATCCACTCTCTGATTTCAATG 120  
Db 710 AGCAGATGGGCAACTGCTGAGAGAGAAAGGTTGAAGATCCACTCTCTGATTTCAATG 769

Qy 121 CAGGAAAAAGTTAGCTGAATATGACCAAGGTAGGCGATTGTGCATGATGGGGCTC 180  
Db 770 CAGGAAAAAGTTAGCTGAATATGACCAAGGTAGGCGATTGTGCATGATGGGGCTC 829

Qy 181 ATCTACGATGGGAGTATTGACGAGTACAAATATGATGAGAAATTTCTACTATCCAAATG 240  
Db 830 ATCTACGATGGGAGTATTGACGAGTACAAATATGATGAGAAATTTCTACTATCCAAATG 889

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
DEFINITION complete cds.



ACCESSION AF039401  
 NID 94009459  
 VERSION AF039401.1 GI:4009459  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 35278)  
 AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.  
 TITLE Genomic cloning, molecular characterization, and functional analysis of human CIC1, the first human member of the family of Ca2+-activated Cl--channel proteins  
 JOURNAL Genomics 54 (2), 200-214 (1998)  
 MEDLINE 99047526  
 REFERENCE 2 (bases 1 to 35278)  
 AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
 FEATURES  
 source 1..35278  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1p22-p31"  
 1..1617  
 /gene="hCIC1"  
 1..33522  
 /gene="hCIC1"  
 767..1265  
 /rpt\_family="I1 transposable element"  
 /rpt\_type="dispersed"  
 1587..1596  
 /gene="hCIC1"  
 /note="putative"  
 1618..1708  
 /gene="hCIC1"  
 /number=1  
 join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)  
 /gene="hCIC1"  
 /product="Calcium-dependent chloride channel-1"  
 1962..2383  
 /db\_xref="GI:4009460"  
 /number=2  
 join(2222..2383,6667..6807,6979..7126,6996..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)  
 /gene="hCIC1"  
 /codon\_start=1  
 /product="calcium-dependent chloride channel-1"  
 /protein\_id="AAC95429.1"  
 /db\_xref="PID:94009460"  
 /translation="MGPRKSSVFILILHLLEGALSNSLIOLNNNGEGIVAVIDNPV  
 EDLTIQIKMVAQASLILPEATGRKRYFNKVALILPETWTKADYRPRKTYKNA  
 DVLVAESTPENGDEFTYEMGNGCKGERIHITPFLAKKLAETGPGKAFVEMAH  
 LRMGVDEYNDEKFLYNSGRIOAVRCASGITVONVVRKCOGSCYTRKCTENVTGL  
 YKGECEFLQSKOTKASIMFAOHVDSIVFCTEONHKEAPKONKONCLNLSQGLFL  
 RDSDEFKTTPMTTOPNPFTSLQIGRIYCLVLDKSGMSATGRNLRLNQAQGLFL  
 LQTVELGSMGVAVTSDSAHVOSELIOINSQSDRLAKRLPAASGGSTISGELISA  
 FTVIRKPYTDCSEIVLITDGEDNTISGCEMVKOSGAILHTVAAGPRASAELEISK  
 MTGILQITASDOVONGLIDAFGLSSGNGAVSQSTOLESGGLTLONSOMNGVIV  
 DSTVGKDTLFTLTWTPQPIILMDPSQKOGFVVDNNTKAAVLIQIGLAKVGTWKY  
 SLOASQDTLITVSRASNATLPITVTSKTKDTSKPPSLVAVNATROGASPTLRA  
 STALIEVNGKTVLELDNGADATKEDVTSRYFTYDNTNGRYKVALGAVN  
 AARRVIPQSGALYIPGIENDEIOMPPREIKMDVVOHKOVCFSTSGSGSPVAS  
 DVNNAIPDLEPPQITDLNMEIHGSLNLTWAPGQDYHGTAKHKTIIIRSTILD

LRDKNESLOYNTTALIPKANSSEVFLFKPENITFENGTLDFIAQAVKDYDKSEI  
 SNIAVSLFIPQTPETPSPDETSPAPCPNHIINSTIGIHILKIMKNIGELQJSLIA  
 exon 6667..6807  
 /gene="hCIC1"  
 /number=3  
 6979..7126  
 /gene="hCIC1"  
 /number=4  
 9696..9801  
 /gene="hCIC1"  
 /number=5  
 15454..15631  
 /gene="hCIC1"  
 /number=6  
 18591..18809  
 /gene="hCIC1"  
 /number=7  
 19783..20010  
 /gene="hCIC1"  
 /number=8  
 22253..22427  
 /gene="hCIC1"  
 /number=9  
 24522..24628  
 /gene="hCIC1"  
 /number=10  
 26644..26859  
 /gene="hCIC1"  
 /number=11  
 27447..27708  
 /gene="hCIC1"  
 /number=12  
 28766..28936  
 /gene="hCIC1"  
 /number=13  
 31840..32079  
 /gene="hCIC1"  
 /number=14  
 32919..33522  
 /gene="hCIC1"  
 /number=15  
 exon 11245 a 6870 c 7230 g 9933 t  
 BASE COUNT  
 ORIGIN  
 Query Match 63.8%; Score 153; DB 11; Length 35278;  
 Best Local Similarity 96.9%; Pred. No. 6.5e-37;  
 Matches 156; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 4 ATGCTGATGTTGTTGGTGTGAGTCTACTCTCCAGGTAATGATGAACCTACACTGAC 63  
 DB 6977 AGCGTGAATGTTCTGTTGTGTGAGTCTACTCTCCAGGTAATGATGAACCTACACTGAC 7036  
 QY 64 AATGGGCACTGTGAGAGAGAGGTGAAGATCCTACTCTCTGATTTCATTCGAG 123  
 DB 7037 AATGGGCAACGTGTGAGAGAGAGGTGAAGATCCTACTCTCTGATTTCATTCGAG 7096  
 QY 124 GAAAAAGTTAGCTGAATGTGACCAAGGATGAGGCAATT 164  
 DB 7097 GAAAAAGTTAGCTGAATGTGACCAAGGATGAGGCAATT 7137  
 RESULT 4  
 AB017156 2937 bp mRNA ROD 04-MAR-1999  
 LOCUS AB017156  
 DEFINITION Mus musculus gob-5 mRNA, complete cds.  
 ACCESSION AB017156  
 NID 93721911  
 VERSION AB017156.1 GI:3721911  
 KEYWORDS gob-5.  
 SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
 ORGANISM Mus musculus

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (sites)
JOURNAL	Komiya,T., Tanigawa,Y. and Hirohashi,S. Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice
MEDLINE	Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999) 99160866
RESEARCHERS	2 (bases 1 to 2937)
AUTHORS	Komiya,T., Tanigawa,Y. and Hirohashi,S.
TITLE	Direct Submission
JOURNAL	Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tooru Komiya, ERYOT, JST, Hirohashi Cell Configuration Project; 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2655, Japan (E-mail: tkom@ccp.jst.go.jp, Tel:81-298-47-7563, Fax:81-298-47-5226)
FEATURES	Location/Oualifiers
Source	1..2937
gene	/organism="Mus musculus" /db_xref="taxon:10090" /cell_type="goblet cell" /dev_stage="adult" /tissue_type="intestine"
CDS	15..2756 /gene="gob-5" /codon_start=1 /protein_id="BA33743.1" /db_xref="PID:d1034712" /db_xref="PID:g93721912" /db_xref="GI:3721912" /translation="MESLCKSPVFLILHLLEGVLSESLIOLNNNGYEGIVAIADHDVPV EDELIQHKIMNTQASPYLEATGKRKYEFNVAILPESKARPEYRPLEFEKNHA DIVLSTSPGNDEPTEHIEAGCEGRKHLPDPLAGKLTYOGPDREFVEENAH FRMVFENYNNDKEKFYLISKRPQAVRCGAALTGNVPRCGSGCITNGKVDIRVIG LYKNCVFPDPHONERKASIFNQNINISVFECTEKHNQAPLDNRNRCALSTWEY IOESDFKQTTPMTAOPPAFSLIQGRVCVLVKSGMLNDRNRNCRSLRE LIQVEQGSWVGWGTAFDSAAYOSELKONGSADRDLIKRLPTVSAGSGISCGLRT AFVIYKKRYPDGSEIYLVLDGEINTISSCDFLKOSGAILHTVALGAAAKELEOLS KMGTSQISDDOVNNGNLVAFAPALSNGNAIKOHSLIESRGVNLONNOMNGSVYA VDSSVRGDTELLTMTTHPTPIFAIMPDSGVONEFIIDTTKVAVLVOGPRAKGFMWK YSIOASSOTELTLTVTSRAASTIPPIVTTPPVNMNTGEPSPVTVASIRGASPIIE ASVIALIESVNGKTVILELDNGAGADTKNDGYSRFIAFDANGRSVAKIMLGQ TSDQRAPAPRNRMAYIDGWIEDDEVARNPPRPTSIVODKQLCFRTSSGGSFVATG LRPAAPIDPLEPCOIDLKASIOGONLVNLTWAPGDYDHGRASNVIIMKSTSIVD LDHFNTSLVQNTTGILPKREASSEIEFEFGMTFGNGDIFIALQAVDKSNLSKEI SNIRVSEVIPAOBPPREDSTPCPDISINSTIPGIHVLIKMMKMLGEMQVTIGLH
BASE COUNT	860 a 718 c 693 g 666 t
ORIGIN	
Query Match	Best Local Similarity 62.0%; Score 148.8; DB 12; Length 2937; Matches 183; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
Oy	1 AAAATGTCGATGTTCTGTTGCTGAGTCTACTCTCCACGTAATGATGACACCCTACACTG 60
Dd	
313	AAAAGCGTGATGTCCTTGATATCAACAACACGAGCCCTTAGCATGATGAGCCCTCACACG 372
Oy	61 AGCAGATGGGCACAGTGGAGAGAAGGGGGAAGAGATCCACCTCACTCCTGATTTCATG 120
Dd	
373	AACATATAGGACATGTGGAGAAAAAGGGGATGAGGATTCACCTGACTCCTGACTTCTTG 432
Oy	121 CAGGAAAAAGTTTACCTGATATGAGACCAACAAGTAGGGCATTTGCTCATGAGTGGGCTC 180
Dd	
433	CAGGAAAGAGCTACTAGCTAGTATGGGCCACAAGACAGAGACGACCTTGTGCATGAGTGGGCTC 492
Oy	181 ATCTACGATGGGAGATTATTCAGCAGATACATAATATGATAGAGAAATTTCTACTATCAATG 240
Dd	
493	ACTTCGATGGGAGATGTTTATGATATCAACAACGACAGAGAAGTTCTACTATATCCAAAG 552

RESULT	5	AF043976	3415 bp	mRNA	PRI	07-APR-1999
LOCUS		AF043976	3415 bp	mRNA		
DEFINITION		Homo sapiens CLCA homolog (hCLCA3) mRNA, complete cds.				
ACCESSION		AF043976				
NID		94572288				
VERSION		AF043976.1	GI:4572288			
KEYWORDS						
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS		1 (bases 1 to 3415)				
TITLE		Gruber,A.D. and Pauli,B.U.				
JOURNAL		Molecular cloning and biochemical characterization of a truncated, secreted member of the human family of Ca2+-activated Cl- channels				
MEDLINE		Biochim. Biophys. Acta 1444 (3), 418-423 (1999)				
REFERENCE		99196715				
AUTHORS		2 (bases 1 to 3415)				
TITLE		Gruber,A.D., Elble,R.C. and Pauli,B.U.				
JOURNAL		Direct Submission				
FEATURES		Submitted (21-JAN-1998) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA				
Source		Location/Qualifiers				
gene		1..3415				
CDS		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		1..3415				
		/gene="hCLCA3"				
		19..807				
		/gene="hCLCA3"				
		/codon_start=1				
		/product="CLCA homolog"				
		/protein_id="AAD3734.1"				
		/db_xref="PID:94572289"				
		/db_xref="GI:4572289"				
		/translation="MVFSLKVFLESLLSPLVKSLSVLTINNNGDGVIAINPSPEE				
		DEKLTQIKENVTPASHLHPLHATQKRAYFRVNSVLIPMTYKSKSEYLIIPKQETDQAD				
		VTVADLYLKGGDDPYTLQYGGCGKGGYIHFTFNFLLNNLNTATGPGKVFHWGWH				
		RMGFEDYNDQPIYISRNRTENTRTSTRITVIMVINECGASGCIARPRRSQTEL				
		YEACTEPIPKRSQAKSEISIVFMQMLDSTVECTEIKHNKEAPNL"				
BASE COUNT		1149 a 625 c 630 g 1011 t				
ORIGIN						
Query Match		41.5% Score 99.6; DB 42; Length 3415;				
Best Local Similarity		64.1% Pred. No. 1.3e-20;				
Matches 150; Conservative		0; Mismatches 84; Indels 0; Gaps 0;				
QY		4 ATGCATGATGTTCTGGTCTGCTGCTACTGCTCCGCCGTAATGATGACCTGACTGAGC 63				
DB		317 AGCCAGATGCTACTGCTGCTGCTACTGCTCTTACCTGGAATACGAGATGATCCCTATACACTTC 376				
QY		64 AGATGGCACTGTGTGGAGAGAAGGGTGAAGAGATCCACCTCACTCTGATTCATTGGAG 123				
DB		377 AATATGACAATGTGTGAATTAAGGACAATATATACATTTTACTCCAAACTTCTGTGA 436				
QY		124 GAAAAAATTAGCTGAATATGACACCAAGATAGGCACTTTGTCCATGAGTGGGCTCATC 183				
DB		437 CTAAATACCTGGCTACCTATGAGGCTCGACAGGTAAGTAATTTGTCCATGGTGGCCCATC 496				
QY		184 TACAGTGGGAGATATTGACGAGTACCAATTAATGATGAGAAATTTCTACTTATCCA 237				
DB		497 TCCGTTGGGAGATATTGATGATGATTAATGTGTGAGACACCACTCATATTTTCCA 550				
RESULT		6				
LOCUS		BTU36445	2984 bp	mRNA	MAN	09-FEB-1996
DEFINITION		Bos taurus calicium-activated chloride channel mRNA, complete cds.				
ACCESSION		U36445				
NID		91184065				
VERSION		U36445.1	GI:1184065			

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE  
CDS

1. 2984  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19. .2730  
/note="epithelial chloride channel"  
/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="pid:g1184066"  
/db\_xref="gi:1184066"  
/translation="MVPLVTLFLTLHPGKSSMYNLINNGDGVIAINPSVPE  
DEKLNQNKEMTEASTLYLPHATKRRVFRNVSILIPMTWKSSEYTLMPKDSYQAE  
VIVANPYLHGDDPTLQYGRGEGQYIHPNLTINLPLYSRGRAPVHAHL  
RWGJEDYNGQDFYISRRNTIETARSTHIGTVIVKCOGSCITRRCRDYDQGL  
YEAKCTIPEKQSTAREIMFOSLHTECTEKTNYHVEANLONKMGKSTVDI  
MNSTDFONTSPTEKAPPTQFSLKSKORVVCVLDKSGMSSEDLFRNOAEL  
FLIOLIEKSLVGVMTFDSVAEIRNNLTIRKIDNYENITANLPOEANGSICIGLK  
AGFOAIIOOSQSTSGSEIILLDGEDNEHSCIEVKOSGVITHIALGPSAKLET  
LSDMTGGRFPAKNDINLTNAFSRISRSRGSITQOTIQLESKALATIEKKVNGTVP  
VSTIGNDTFFVWWTIKRPILLDDPKKKYKTSDFEKDKNISARLRIPGIETG  
TWYLSLNHNSPOLITVTVTRARSPTTPVTAHMSONTAHPSPVIVAYOSOG  
FLPVGLINVTALIEREDGHVTLMDNGAGADYVKNQGISRYFTDRNGRVSILKY  
HAARANNTRARLSRQPNKALYIPYIENGKIILNPPREYKDKLAKAIEDPSLTS  
GGSTVSGAPGNHPSVLPNKITIDLEAKFKEDH1QLSWTAPANVLDGKANSYIIRI  
KGSFLDQKDFDNALVNTSLKPKRAGSDENEFKPPFRLENGTNFIYVQAINEA  
NLTSVSNIAQAIKRIIPMPEDSVPLGTRISAINIATPALMIISIV"

BASE COUNT 1027 a 583 c 561 g 813 t  
ORIGIN

Query Match 40.8%; Score 98; DB 3; Length 2984;  
Best Local Similarity 63.7%; Pred. No. 4,1e-20;  
Matches 149; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 4 ATGCGATGTTTGTGTTGCTGAGTCTACTCTCCAGGTAATGATGACCTACACTGAC 63  
DB 317 AGCGAGAACTCACTAGTCTAATCTTACTTAAACATGAGATGATCCCTTACTTC 376  
QY 64 AATGGGCACTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 123  
DB 377 AATATGGAAGATGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 436  
QY 124 GAAAAAGTTAGCTGAATATGACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 183  
DB 437 CTAAATTAATTTGCCATATTCATGATGGGTGAGGAGGAGGAGGAGGAGGAG 496  
QY 184 TACGATGGGAGATATTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 237  
DB 497 TCCGATGGGAGATATTGATGATATACGAGGAGGAGGAGGAGGAGGAGGAGGAG 550

AF001261  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE  
CDS

1. 3317  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="1"  
/tissue\_type="lung"  
63. .2780  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86529.1"  
/db\_xref="pid:g2623763"  
/db\_xref="gi:2623763"  
/translation="MVLCNLYLFLTLHPGKSSMYNLINNGDGVIAINPSVPE  
DEKLNQNKEMTEASTLYLPHATKRRVFRNVSILIPMTWKSSEYTLMPKDSYQAE  
VIVANPYLHGDDPTLQYGRGEGQYIHPNLTINLPLYSRGRAPVHAHL  
RWGJEDYNGQDFYISRRNTIETARSTHIGTVIVKCOGSCITRRCRDYDQGL  
YEAKCTIPEKQSTAREIMFOSLHTECTEKTNYHVEANLONKMGKSTVDI  
MNSTDFONTSPTEKAPPTQFSLKSKORVVCVLDKSGMSSEDLFRNOAEL  
FLIOLIEKSLVGVMTFDSVAEIRNNLTIRKIDNYENITANLPOEANGSICIGLK  
AGFOAIIOOSQSTSGSEIILLDGEDNEHSCIEVKOSGVITHIALGPSAKLET  
LSDMTGGRFPAKNDINLTNAFSRISRSRGSITQOTIQLESKALATIEKKVNGTVP  
VSTIGNDTFFVWWTIKRPILLDDPKKKYKTSDFEKDKNISARLRIPGIETG  
TWYLSLNHNSPOLITVTVTRARSPTTPVTAHMSONTAHPSPVIVAYOSOG  
FLPVGLINVTALIEREDGHVTLMDNGAGADYVKNQGISRYFTDRNGRVSILKY  
HAARANNTRARLSRQPNKALYIPYIENGKIILNPPREYKDKLAKAIEDPSLTS  
GGSTVSGAPGNHPSVLPNKITIDLEAKFKEDH1QLSWTAPANVLDGKANSYIIRI  
KGSFLDQKDFDNALVNTSLKPKRAGSDENEFKPPFRLENGTNFIYVQAINEA  
NLTSVSNIAQAIKRIIPMPEDSVPLGTRISAINIATPALMIISIV"

BASE COUNT 1084 a 676 c 634 g 923 t  
ORIGIN

Query Match 39.5%; Score 94.8; DB 3; Length 3317;  
Best Local Similarity 62.8%; Pred. No. 4e-19;  
Matches 147; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 4 ATGCGATGTTTGTGTTGCTGAGTCTACTCTCCAGGTAATGATGACCTACACTGAC 63  
DB 361 AGCGAGAACTCACTAGTCTAATCTTACTTAAACATGAGATGATCCCTTACTTC 420  
QY 64 AATGGGCACTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 123  
DB 421 AATATGGAAGATGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480  
QY 124 GAAAAAGTTAGCTGAATATGACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 183

Db 481 CTAATTAATTCACATCTATGGGCTCCCGACAGATATTTCATGATGCGCCCATC 540

Qy 184 TACGATGGGAGTATTGACGAGTACATATGATGAGAAATTCATCTATCA 237

Db 541 TCCGCTGGGAGTATTGATGAGTAAATGTGACACGACATCTATATTCCA 594

RESULT 8

AF001262 3288 bp mRNA MAM 19-NOV-1997

LOCUS Bos taurus clone 2 endothelial adhesion molecule Lu-ECAM-1 mRNA, complete cds.

AF001262

AF001262.1 GI:2623764

VERSION 92823764

KEYWORDS

SOURCE Bos taurus.

ORGANISM Bos taurus.

REFERENCE 1 (bases 1 to 3288)

AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.

TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3288)

AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.

TITLE Direct Submission

JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA

FEATURES

source Location/Qualifiers

1. .3288

/organism="Bos taurus"

/db\_xref="taxon:9913"

/cell\_type="aortic endothelial cells"

/clone="2"

/tissue\_type="lung"

59. .2443

/note="endothelial adhesion molecule; chloride channel"

/codon\_start=1

/product="Lu-ECAM-1"

/protein\_id="AAB86530.1"

/db\_xref="PID:92623765"

/db\_xref="GI:2623765"

/translation="MVLCLNVILFLTLHPGKSSMVLINNGYDGIIVAINPSVE DEKLINIKEMVTEASTYLEPHATKRVRVFNVSILIPMTKRSSEYFLPROESYDAD VIVANPLYKGDPPYTIQYRCGCEKRYIHFTPNFLTNHNYGSRGVFVEMAH RRGIFEDINVDQPTISRKNTIATRCSTHITGINVFKCPGSCITSLCRDSQIG LYENKCFLEPKSQTAKEISIMFPLSHSVETECTHTNEAPLNOKMCKNGSTWY LMSVDFONTSPMTENMPHPHPSLSKSKORVCLVLDKSGMSADRLFQNMQAE LYLIOVLEKSGLVGMYTFDVAEIOHNLRTIDNVYOKITAKLPVANGSTICRGL KAGQALIHSDOSTSGSEIILTDGDNENISCFEDVYKSGAIIHITAIQPSAKKEE TLSMNGGYRFRANKDITGILTNAPFSISSSGSIITQALDESAKITGRKRVNTY PVDSTVGNDTFEYVVTWITQPELVDPKQKRTISDFEKLNISARLQIPGIAET GTWYISLNNHASSOMLTVVTRARSPITPVIATAHMSQHTAHYSPMIVAYQVSO GFLPVLGISVIAIETEDGHQVLELMDNGAGDVTAKNDGYSRYFDYDNGRYSIAK VHAQARNNTARLNLROPNKVLVPGVYENGKIIINPPEVYKDLAKAKIEDSRILT SGGSFTVSGARPPGNHPSVPPSKITDLEAKFKEDYIQLSWTAPGVLDKRGKSSFP

BASE COUNT 1043 a 707 c 645 g 893 t

ORIGIN

Query Match 39.5%; Score 94.8; DB 3; Length 3288;

Best Local Similarity 62.8%; Pred. No. 4e-19;

Matches 147; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 4 ATGCTGATGTTCTGCTGAGTCTACTCTCCAGCATGATGAGAACCTACACTGAGC 63

Db 357 AGCGAGATGTCATAGTCTATCCCTAATAAATATGAGATGATCCCTATACACTTC 416

Qy 64 AGATGGCACTGTGGAGACGAGAGGTGAAGATCCACCTACTCTGATTTCATGCG 123

Db 417 AATATGGAAGGTGGAGAAAAAGAAAAATATATACATTTCATCCAAACTCTTGTTGA 476

Qy 124 GAAAAAGTATGCGAATATGAGACCAAGGTAGGGCAATTGTCCATGAGGGCTCATC 183

Db 477 CTAATTAATTCACATCTATGGGCTCCCGACAGATATTTCATGATGCGCCCATC 536

Qy 184 TACGATGGGAGTATTGACGAGTACATATGATGAGAAATTCATCTATCA 237

Db 537 TCCGCTGGGAGTATTGATGAGTAAATGTGACACGACATCTATATTCCA 590

RESULT 9

AF001263 2820 bp mRNA MAM 19-NOV-1997

LOCUS Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA, complete cds.

AF001263

AF001263.1 GI:2623766

VERSION 92823766

KEYWORDS

SOURCE Bos taurus.

ORGANISM Bos taurus.

REFERENCE 1 (bases 1 to 2820)

AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.

TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2820)

AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.

TITLE Direct Submission

JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA

FEATURES

source Location/Qualifiers

1. .2820

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone="3"

/tissue\_type="lung"

/cell\_type="aortic endothelial cells"

194. .2656

/note="endothelial adhesion molecule; chloride channel"

/codon\_start=1

/product="Lu-ECAM-1"

/protein\_id="AAB86531.1"

/db\_xref="PID:92623767"

/db\_xref="GI:2623767"

/translation="MVLCLNVILFLTLHPGKSSMVLINNGYDGIIVAINPSVE DEKLINIKEMVTEASTYLEPHATKRVRVFNVSILIPMTKRSSEYFLPROESYDAD VIVANPLYKGDPPYTIQYRCGCEKRYIHFTPNFLTNHNYGSRGVFVEMAH RRGIFEDINVDQPTISRKNTIATRCSTHITGINVFKCPGSCITSLCRDSQIG LYENKCFLEPKSQTAKEISIMFPLSHSVETECTHTNEAPLNOKMCKNGSTWY LMSVDFONTSPMTENMPHPHPSLSKSKORVCLVLDKSGMSADRLFQNMQAE LYLIOVLEKSGLVGMYTFDVAEIOHNLRTIDNVYOKITAKLPVANGSTICRGL KAGQALIHSDOSTSGSEIILTDGDNENISCFEDVYKSGAIIHITAIQPSAKKEE TLSMNGGYRFRANKDITGILTNAPFSISSSGSIITQALDESAKITGRKRVNTY PVDSTVGNDTFEYVVTWITQPELVDPKQKRTISDFEKLNISARLQIPGIAET GTWYISLNNHASSOMLTVVTRARSPITPVIATAHMSQHTAHYSPMIVAYQVSO GFLPVLGISVIAIETEDGHQVLELMDNGAGDVTAKNDGYSRYFDYDNGRYSIAK VHAQARNNTARLNLROPNKVLVPGVYENGKIIINPPEVYKDLAKAKIEDSRILT SGGSFTVSGARPPGNHPSVPPSKITDLEAKFKEDYIQLSWTAPGVLDKRGKSSFP MSFESHQVAIVLELQLOHOSFO

BASE COUNT 940 a 565 c 560 g 755 t

ORIGIN

Query Match 39.5%; Score 94.8; DB 3; Length 2820;



[illegible]

BASE COUNT ORIGIN	1096 a	777 c	736 g	862 t
----------------------	--------	-------	-------	-------

Query Match	39.4%	Score	94.6	DB	12	Length	3471
Best Local Similarity	62.4%	Pred.	No. 4,6e-19				
Matches	148	Conservative	0	Mismatches	89	Indels	0
				Gaps	0		

  

QY	1	AAATGTCGATGTTCTGGTTCCTGAGTCTACTCCTCCAGTAATGATGATACCTACACTG	60
DB	342	ACAAGGACGAGCTATAGTTCGGGATCCTCAGTCGACATGAGAGACGCCCTACACCC	401
QY	61	AGCGATGGCCAACTGTGTGAGAGAAAGGTGAAAGATCCACTCACTCTTATTTCAATG	120
DB	402	TTCATATGACAGAGTGTGGGACAGAGCAGTCAATCACTTCACTCCAAACTCTCCAC	461
QY	121	CAGGAAAAAGTTAGCTCAATATGACACCAAGGTAGAGGCAATTTGTCATGAGTGGCTC	180
DB	462	TCACGTAACTTGCGTATCTACGAGACCCCGACGACAGTCTTTGTCCATGATGAGGCC	521

Db 522 ATCTCCGGTGGGAGTATTGATGAGTATTAACGTGGACCAACCCTTTCTACATGTCTA 578

LOCUS	AB026833	3604 bp	PR1	26-MAY-1998
DEFINITION	Homo sapiens mRNA for chloride channel protein, complete cds.			
ACCESSION	AB026833			
NID	94887600			

```

VERSION      AB026833.1 GI:4887600
KEYWORDS     chloride channel protein.
SOURCE       Homo sapiens adult corneal epithelium cDNA to mRNA,
              clone_lhb:lambda Zap clone:lambda 13802.
              accession:U00000

```

REFERENCE  
AUTHORS  
TITLE

Eutheria; Primates; Carnivora; Hominoidea; Homo.  
1 (sites)  
Itoh, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.  
Isolation and characterization of chloride channel in human corneal

REFERENCE  
2 (bases 1 to 3604)  
AUTHORS  
TITLE  
Itoh, R., Kawamoto, S. and Okubo, K.  
Direct Submission  
Submitted (28-APR-1999) to the DDBJ/EMBL/Genbank databases. Reiko Itoh, Institute for Molecular and Cellular Biology, Osaka

## FEATURES

```

/cclone_lib="lambda zap"
/dev_stage="adult"
/map="ip32"
/tissue_type="corneal epithelium"
42.2873
/codon_start=1
CDS

```



CDS  
2904..3653  
/gene="MJ0768"  
/note="similar to GB:L77117 PID:1592005 percent identity: 40.70; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAB98763.1"  
/db\_xref="PID:g1499588"  
/db\_xref="GI:1499588"  
/translation="MIKKRVEVYIGLELPFGSGEQINLSLIAQPIEDGDIIVIA  
ETLSKLEGGVIDDKIIPSKAEIAELAKTGDKDPVQVILDEAKETKVKKNLIE  
TKHGVCANSGVDSNTYKIKIILPKNPDESAREIKREIEKTRGVCVILSDSGRP  
FRKAVGAIIVSGIILAMDREKEDLFGRELKTTEVAIADELASMANVVGDEDEGI  
PVILRGANVPFGNGKGRDLIRPEEDVFRN"  
complement(3793..4302)  
/gene="MJ0769"  
complement(3793..4302)  
/gene="MJ0769"  
/note="hypothetical protein; identified by Genemark;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="M. jannaschii predicted coding region MJ0769"  
/protein\_id="AAB98774.1"  
/db\_xref="PID:g1499589"  
/db\_xref="GI:1499589"  
/translation="MIEIKISKIPRMDEINKIVLREKDLVLLKPSVSEHPKMAK  
LEYLKKGFIIEVNAKGRKRYDDEKRIHEIIEGYSVREIGNLIGKSTVMD  
YAKDCKELERFRKLIWVEYREYLINKGKSPSLOVLFLELNAVDEDEKAKKILE  
DIRKVEFR"  
complement(4295..4966)  
/gene="MJ0770"  
complement(4295..4966)  
/gene="MJ0770"  
/note="hypothetical protein; identified by Genemark;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="M. jannaschii predicted coding region MJ0770"  
/protein\_id="AAB98775.1"  
/db\_xref="PID:g1499590"  
/db\_xref="GI:1499590"  
/translation="MIMIDWNLKSPSESIKHSCTWMDKIDYKOTYKFEKDLKLN  
KELLYEDDYKRIAYLITFLFQLRNGCRLWEIAIMINAINIDNLMMNERITVVRQ  
KRKDWERELIIPKCIKEDIEWVDFLIDKIDELKIDELKAKKIVKRFGAM  
LYKNYGINTHSLRAYVYTLGEGHIGPAQVLAKITKKNINIEITYQSRLAKETIKNI  
GDLDD"  
5055..5747  
/gene="MJ0771"  
5055..5747  
/gene="MJ0771"  
/note="similar to GB:M59301 SP:P21639 PID:151174  
PID:1926450 percent identity: 27.71; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="cobalamin biosynthesis precorin-2  
methyltransferase (cbiL)"  
/protein\_id="AAB98764.1"  
/db\_xref="PID:g1592306"  
/db\_xref="GI:1592306"  
/translation="MNKLKRVYGVGVGVDKLLLTAKALEVLKKVDKIRVPSKKKQ  
KSIAYEIKDVDDKNEIELFPMIKDKERLAKTWNALEVKLEDEVALITIGDPT  
LSITFVYWKLLRGVGEIVNGISSIFASAAALNPLVGDGKLCILPGKDLERY  
IDEDDTIIMKTKLNKLSVKNRDYIIGLVKRAFEDKVVYIGLDEINPEFND  
YLSAIIKRRK"  
5744..6805  
/gene="MJ0772"  
5744..6805  
/gene="MJ0772"  
/note="hypothetical protein; identified by Genemark;  
putative"

Query Match 15.0% Score 36; DB 2; Length 19436;  
Best Local Similarity 56.9% Pred. No. 0.51; Gaps 0;  
Matches 66; Conservative 0; Mismatches 50; Indels 0;  
gene  
81 GAGAGGGGTGAAGAGCACCCTCCATTCGATTCATGCAGAAAGAAATAGCGAA 140  
Db 18940 GAAAGGGCTGAATAGCTGTGCTAAATGTGATGTGTATCAAGAAAATTTTGAT 18881  
QY 141 TATGACCAAGAGTAGGCAATTTGCCATGATGAGTGGGCTATCAGATGGGAGT 196  
Db 18880 ATTGGACCAAGATGAGGCGATTTGACCATGACCAAGAAATTAAGATGTAGAGT 18825  
RESULT 15  
MEPGK1 1251 bp mRNA MAM 29-JUL-1996  
LOCUS M.eugenii mRNA fragment for phosphoglycerate kinase.  
ACCESSION X64296  
NID 91149  
VERSION X64296.1 GI:1149  
KEYWORDS pgk gene; phosphoglycerate kinase.  
SOURCE tammar wallaby.  
ORGANISM Macropus eugenii  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Metatheria;  
Diprotodontia; Macropodidae; Macropus.  
REFERENCE 1 (bases 1 to 1251)  
AUTHORS Zehafi-Feferman,T.  
TITLE Direct Submission  
JOURNALS Submitted (24-JAN-1992) T. Zehafi-Feferman, School of Biological  
Sciences, Macquarie University, NSW 2109, AUSTRALIA  
REFERENCE 2 (bases 1 to 1251)  
AUTHORS Zehavi-Feferman,R. and Cooper,D.W.  
TITLE PCR derived cDNA clones for x-linked phosphoglycerate kinase-1 in a  
marsupial, the tammar wallaby (Macropus eugenii)  
JOURNALS Biochem. Biophys. Res. Commun. 187 (1), 26-31 (1992)  
MEDLINE 92392333  
FEATURES  
source Location/Qualifiers  
gene 1..1251  
/organism="Macropus eugenii"  
/db\_xref="taxon:9315"  
/chromosome="X"  
1..1251



CDS

```

/gene="PGK-1"
<!. .1251
/gene="PGK-1"
/codon_start=1
/product="phosphoglycerate kinase"
/protein_id="CAA4574.1"
/db_xref="PDB:q1150"
/db_xref="GI:1150"
/db_xref="SWISS-PROT:P29408"
/translation="SLSKLTLDKVDLKGKRYIMRVDFNVPMKNEITNMORIKALP
SINYCLDNGAKSVILMSHGRPDGVPMPDKYSLPEVAELKSLGKDYFLKDCVPE
VEOACANPNKSIITILENRFHVEEKGKDSGNKIKAEPAKMEAFQASIKLDVY
VNDAGTARAHSSMYGVNLPKACGFLMKKELTFPAKALDSPEPFLLIIGAKAYAD
KIQLINMLDKVNEIIGGGMGFTFLKVLNNEIGTSLDEEGAKIVKDLMAKAEKG
VKITLPVDEVTADKFDENAKTGQATLASGIPAGMGLDGPSSKRYEYVYVMAKQIV
WNGPVGFEMEEFARGTKELMNNVVEATFRGCITIIIGGDIATCCAKWNTEDKVSHSV
TGGGASLELLEGGKVLPGVSTLNNV"

```

BASE COUNT 333 a 276 c 353 g 289 t  
ORIGIN

Query Match 14.4% Score 34.6; DB 3; Length 1251;

Best Local Similarity 61.8% Pred. 1.1; Mismatches 34; Indels 0; Gaps 0;

Matches	55;	Conservative	0;	Mismatches	34;	Indels	0;	Gaps	0;
QY	56	CACTGACGACATGGGCACTGTGGAGAGAGGGGTGAAGGATCCACCTCCTGATTT	115						
DB	795	CAAGGACCTGATGGCCAGGCTGAGAAACGGCGTCAAGATCACGCTGCTGTGACTT	854						
QY	116	CATTGACGAGAAAAAGTTAGCTGAATATG	144						
DB	855	CGTCACGACGACAGTGTGATGAATAATG	883						

Search completed: August 5, 1999, 23:10:15  
Job time: 7498 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:24 ; Search time 3059.95 Seconds

(Without alignments)  
154.711 Million cell updates/sec

Title: US-09-049-696-3

Perfect score: 240  
Sequence: 1 AAAATGCTGATGTTCTGTT.....GAAATTCCTACTATCCAAATG 240

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	20	8.3	430	25	N70553	N70553 za83e12.s1
2	20	8.3	420	32	AA342991	AA342991 EST48700
3	20	8.3	388	37	AA714919	AA714919 nv49p03.r
4	19	7.9	557	22	H10546	H10546 ym04d01.r1
5	19	7.9	274	31	AA298669	AA298669 EST114502
6	19	7.9	480	35	AA581198	AA581198 nd38c07.r
7	19	7.9	415	37	AA669224	AA669224 ac16b04.s
8	18	7.5	518	26	z71911	z71911 CHEST809 GO
9	18	7.5	449	41	AI022693	AI022693 ow67c09.x
10	17	7.1	572	21	T62858	T62858 yco3b08.s1
11	17	7.1	438	22	R42419	R42419 yf92e04.s1
12	17	7.1	388	28	AA112392	AA112392 zn68g12.s
13	17	7.1	547	35	AA546137	AA546137 VK62b04.r
14	17	7.1	471	46	AI459786	AI459786 ap17e06.x
15	16	6.7	361	5	AF124520	AF124520 Homo sapi
16	16	6.7	42	20	T17572	T17572 mps v338.Th
17	16	6.7	400	21	F14087	F14087 ATTS5009 Ve
18	16	6.7	394	22	R47107	R47107 y622 Rat in
19	16	6.7	456	22	R61013	R61013 yh05a06.r1
20	16	6.7	284	23	H33607	H33607 EST109760 R
21	16	6.7	425	24	H96411	H96411 yw61e06.s1
22	16	6.7	292	26	W18502	W18502 mb88b07.r1
23	16	6.7	410	28	AA098408	AA098408 mol2f08.r
24	16	6.7	464	28	D44880	D44880 HDMSUP343
25	16	6.7	335	29	AA183650	AA183650 mt20a03.r
26	16	6.7	454	30	AA211295	AA211295 zq87g01.r
27	16	6.7	439	30	AA211296	AA211296 zq87g01.s
28	16	6.7	310	32	AA332508	AA332508 EST36479
29	16	6.7	340	32	AA346272	AA346272 EST52476
30	16	6.7	324	32	AA355660	AA355660 EST64367
31	16	6.7	521	34	AA454314	AA454314 MBAPCXB08
32	16	6.7	494	34	AA511177	AA511177 V938d01.r
33	16	6.7	528	35	AA547106	AA547106 vk72h10.s
34	16	6.7	624	36	C76682	C76682 C76682 Mous
35	16	6.7	347	37	AA683045	AA683045 ae89f03.s
36	16	6.7	404	39	AA841076	AA841076 MB3D6A3C
37	16	6.7	473	39	AA847714	AA847714 oe20d01.s
38	16	6.7	304	39	AA852121	AA852121 NHTBcae10
39	16	6.7	380	39	AA856819	AA856819 cd83a08.s
40	16	6.7	527	40	AA839923	AA839923 vq64c05.s
41	16	6.7	801	41	AI048556	AI048556 ud61c08.y
42	16	6.7	776	43	AI186064	AI186064 qe50h05.x
43	16	6.7	450	45	AI353903	AI353903 zeh1h37.s
44	16	6.7	431	47	AI480552	AI480552 vg38d06.x
45	16	6.7	263	51	AI735118	AI735118 as76b02.x

#### ALIGNMENTS

RESULT 1  
LOCUS N70553 430 bp mRNA EST 14-MAR-1996  
DEFINITION za83e12.s1 Soares\_fetal\_lung\_NDHL19W Homo sapiens CDNA clone  
IMAGE:299182 3', mRNA sequence.  
ACCESSION N70553  
NID G1227133  
VERSION N70553.1 G1:1227133

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS 1 (bases 1 to 430)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holtman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maria, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
On Apr 14, 1993 this sequence version replaced gi:785691.  
JOURNAL  
COMMENT

TITLE  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: ml3 -40 forward  
High quality sequence stop: 344.  
Location/Qualifiers  
1. 430  
/organism="Homo sapiens"  
/db\_xref="GDB:1244106"  
/db\_xref="taxon:9606"  
/clone\_image="299182"  
/clone\_lib="Soares\_fetal\_lung\_NDH19W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: pTZ19 (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCATCTGAAGTGGAGCGGCCCAATTTTCTTTTCTTTT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pTZ19 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NDH19W."

BASE COUNT 167 a 56 c 65 g 140 t 2 others  
ORIGIN

Query Match 8.3%; Score 20; DB 25; Length 430;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 TGTGGAGAGAGGCTGAAG 94  
|||||

Db 217 TGTGGAGAGAGGCTGAAG 236

RESULT 2  
AA342991 420 bp mRNA EST 21-APR-1997  
DEFINITION ES18700 Fetal spleen Homo sapiens cDNA 3' end, mRNA sequence.  
ACCESSION AA342991  
NID 91995292  
VERSION AA342991.1 GI:1995292  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS 1 (bases 1 to 420)  
Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C. and Venter, J.C.

TITLE  
Rapid cDNA sequencing (expressed sequence tags) from a  
directionally cloned human infant brain cDNA library  
JOURNAL Nature Genet. 4, 373-380 (1993)  
MEDLINE 94004965  
COMMENT On May 8, 1995 this sequence version replaced gi:801175.  
Other ESTs: THC147697  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13-21.  
Location/Qualifiers  
1. 420  
/organism="Homo sapiens"  
/db\_xref="ATCC (lnhost):144542"  
/db\_xref="taxon:9606"  
/clone\_lib="Fetal spleen"  
/dev\_stage="fetus"  
/note="Organ: spleen; Vector: pBluescript SK-; Site 1:  
EcoRI; Site 2: XhoI"

BASE COUNT 146 a 58 c 67 g 143 t 6 others  
ORIGIN

Query Match 8.3%; Score 20; DB 32; Length 420;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 TGTGGAGAGAGGCTGAAG 94  
|||||

Db 133 TGTGGAGAGAGGCTGAAG 152

RESULT 3  
AA714919 388 bp mRNA EST 18-FEB-1998  
DEFINITION nv49b03.t1 NCI-CGAP\_Ew1 Homo sapiens cDNA clone IMAGE:1233101, mRNA  
sequence.  
ACCESSION AA714919  
NID 92727193  
VERSION AA714919.1 GI:2727193  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS 1 (bases 1 to 388)  
NCI-CGAP <http://www.nci.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1404660.  
JOURNAL  
COMMENT

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.  
cDNA Library Preparation: David B. Kitzman, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/bbrp/image/image.html](http://www.bio.lnl.gov/bbrp/image/image.html)  
Insert Length: 505 Std Error: 0.00  
Seq primer: -28ml3 rev1 ET from Amersham.

FEATURES  
source  
Location/Qualifiers  
1. .388  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1233101"  
/clone\_1db="NCI CGAP Ew1"  
/tissue-type="Ewing's sarcoma"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; mRNA made from Ewing's sarcoma, cDNA made by oligo-dT priming. Non-directionally cloned. Size selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT  
ORIGIN  
140 a 52 c 63 g 133 t

Query Match  
Best Local Similarity 100.0%; Score 20; DB 37; Length 388;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 TGTGAGAGAGAGGTGAAG 94  
|||||  
Db 254 TGTGAGAGAGAGGTGAAG 273

RESULT 4  
LOCUS  
H10546  
DEFINITION  
H10546 557 bp mRNA EST 23-JUN-1995  
Y004601.r1 Soares infant brain INIB Homo sapiens cDNA clone  
IMAGE:46959 5' similar to SP:KPIM\_HUMAN P11309 P1M-1 PROTO-ONCOGENE  
SPRINE/THREONINE-PROTEIN KINASE ;, mRNA sequence.  
H10546  
9675368  
H10546.1 GI:875368  
EST.  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS  
1 (bases 1 to 557)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Ten, F.,  
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
On May 10, 1995 this sequence version replaced gi:805969.

TITLE  
JOURNAL  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
Insert Size: 1866  
High quality sequence stops: 400  
Source: IMAGE Consortium, LML  
This clone is available royalty-free through LML; contact the  
IMAGE Consortium (info@image.lml.gov) for further information.  
Insert Length: 1866 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stops: 400.  
Location/Qualifiers  
1. .557  
/organism="Homo sapiens"  
/db\_xref="GDB:419500"  
/db\_xref="taxon:9606"  
/clone="IMAGE:46959"  
/clone\_1db="Soares infant brain INIB"  
/sex="female"  
/dev\_stage="73 days post natal"

/lab\_host="DH10B (ampicillin resistant)."  
/note="Organ: whole brain. Vector: lacmid BA; Site\_1: Not  
I; Site\_2: Hind III; 1st strand cDNA was primed with a Not  
I - oligo(dT) primer [5',  
AAGCTGAGAGATTCGGCGCCGACGATATTTTATTTTATTTT 3'];  
double-stranded cDNA was ligated to Hind III adaptors  
(Pharmacia), digested with Not I and directionally cloned  
into the Not I and Hind III sites of the lacmid BA vector.  
Library went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  
ORIGIN  
108 a 150 c 152 g 141 t 6 others

Query Match  
Best Local Similarity 100.0%; Score 19; DB 22; Length 557;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 ATGATGAACCTACACGTA 61  
|||||  
Db 351 ATGATGAACCTACACGTA 369

RESULT 5  
LOCUS  
AA298669  
DEFINITION  
AA298669 274 bp mRNA EST 18-APR-1997  
EST114502 Pancreas tumor I Homo sapiens cDNA 5' end similar to  
similar to proto-oncogene pim-1, mRNA sequence.  
AA298669  
G1951236  
NID  
AA298669.1 GI:1951236  
EST.  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS  
1 (bases 1 to 274)  
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,  
Bult, C.D., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,  
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, Wai, C.,  
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,  
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geochagen, N.S.,  
Giordano, J., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,  
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,  
Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y.,  
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,  
He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,  
Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, R.,  
Raymond, L., Wei, Y.F., Wang, Y., Xu, C., Yu, G.L., Ruben, S.M.,  
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,  
Fraser, C.M. and Venter, J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl.), 3-174 (1995)  
96026280  
JOURNAL  
MEDLINE  
COMMENT  
On Sep 12, 1996 this sequence version replaced gi:1393028.  
Other ESTs: TRC189235  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerl@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/cdb/ngi/ngi.html>)  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .274

/organism="Homo sapiens"  
/db\_xref="ATCC (Innoc):190362"  
/db\_xref="taxon:9606"  
/map="19p12-p13.1"  
/clone\_lib="Pancreas tumor 1"  
/dev\_stage="adult"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"  
BASE COUNT 58 a 78 c 71 g 65 t 2 others  
ORIGIN

Query Match 7.9%; Score 19; DB 31; Length 274;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 ATGATGAACCTTACACTGA 61  
|||||  
Db 218 ATGATGAACCTTACACTGA 236

RESULT 6  
AA581198 480 bp mRNA EST 05-JAN-1998  
LOCUS nd38c07.r1 NCI CGAP Col Homo sapiens cDNA clone IMAGE:802572.5  
DEFINITION similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.  
; mRNA sequence.  
ACCESSION AA581198  
NID 92358970  
VERSION AA581198.1 GI:2358970  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 480)  
NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 9, 1995 this sequence version replaced gi:803001.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies Inc., David Kitzman,  
Ph.D.  
CDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbp/image/image.html  
Seg primer: -28m3 rev1 ET from Amersham  
High quality sequence stop: 405.  
Location/Qualifiers  
1. 480  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:802572"  
/clone\_lib="NCI CGAP Col"  
/tissue\_type="bulk tumor"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: PCMV-SPORT2; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dr."  
BASE COUNT 168 a 101 c 97 g 114 t  
ORIGIN

Query Match 7.9%; Score 19; DB 35; Length 480;  
Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 222 AATTTCTACTATCCCATG 240  
|||||  
Db 1 AATTTCTACTATCCCATG 19

RESULT 7  
AA669224 415 bp mRNA EST 20-NOV-1997  
LOCUS ac16b04.s1 Stratagene ovary (#937217) Homo sapiens cDNA clone  
DEFINITION IMAGE:856591.3, similar to TR:G765068 G765068  
PROTEIN-SERINE/THREONINE KINASE. [2] TR:G765067; mRNA sequence.  
ACCESSION AA669224  
NID 92630723  
VERSION AA669224.1 GI:2630723  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 415)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krieman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F.,  
Theisling, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Nov 6, 1997 this sequence version replaced gi:932378.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 409.  
Location/Qualifiers  
1. 415  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:856591"  
/clone\_lib="Stratagene ovary (#937217)"  
/sex="female"  
/dev\_stage="49 year old"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: ovary; Vector: Bluescript SK; Site\_1: EcoRI;  
Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dr.  
Total ovary tissue, normal, Caucasian. Average insert  
size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'  
GAATTCGCGCAGCAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTCTTTTCTTTT 3' "

BASE COUNT 88 a 102 c 101 g 124 t  
ORIGIN

Query Match 7.9%; Score 19; DB 37; Length 415;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 ATGATGAACCTTACACTGA 61  
|||||  
Db 190 ATGATGAACCTTACACTGA 208

RESULT 8  
271911/c 518 bp mRNA EST 13-NOV-1996  
LOCUS CHEST809 Goat mammary gland Capra hircus cDNA clone EST8-09, mRNA  
DEFINITION

```

sequence.
ACCESSION 271911
NID 91292836
VERSION 271911.1 GI:1292836
KEYWORDS EST.
SOURCE goat.
ORGANISM Capra hircus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Caprinae; Capra.
REFERENCE 1 (bases 1 to 518)
AUTHORS Le Provost, F., Lepoint, A. and Martin, P.
TITLE A survey of the goat genome transcribed in the lactating mammary
gland
JOURNAL Mamm. Genome 7 (9), 657-666 (1996)
MEDLINE 96359152
COMMENT On Apr 14, 1993 this sequence version replaced gi:785308.

Contact: Le Provost, F.
Laboratoire de Genetique Biochimique
INRA-CRU
78352 Jouy-en-Josas Cedex, FRANCE.

FEATURES
source
1. 518
/organism="Capra hircus"
/db_xref="taxon:9925"
/clone_lib="EST8-09"
/clone_1lb="Goat mammary gland"
/note="Mammary gland of Capra hircus (goat) "

BASE COUNT 161 a 95 c 94 g 159 t 9 others
ORIGIN

Query Match 7.5%; Score 18; DB 26; Length 518;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AATGCTGATGTTCTGTT 20
|||||
Db 61 AATGCTGATGTTCTGTT 44

RESULT 9
LOCUS AI022693 449 bp mRNA EST 28-AUG-1998
DEFINITION ow67609.x1 Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens CDNA
ACCESSION AI022693
NID 93237934
VERSION AI022693.1 GI:3237934
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 449)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR Gene Index
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2284689.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 2338 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 357.
location/Qualifiers
1. 449
/organism="Homo sapiens"

sequence.
ACCESSION 271911
NID 91292836
VERSION 271911.1 GI:1292836
KEYWORDS EST.
SOURCE goat.
ORGANISM Capra hircus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Caprinae; Capra.
REFERENCE 1 (bases 1 to 518)
AUTHORS Le Provost, F., Lepoint, A. and Martin, P.
TITLE A survey of the goat genome transcribed in the lactating mammary
gland
JOURNAL Mamm. Genome 7 (9), 657-666 (1996)
MEDLINE 96359152
COMMENT On Apr 14, 1993 this sequence version replaced gi:785308.

Contact: Le Provost, F.
Laboratoire de Genetique Biochimique
INRA-CRU
78352 Jouy-en-Josas Cedex, FRANCE.

FEATURES
source
1. 518
/organism="Capra hircus"
/db_xref="taxon:9925"
/clone_lib="EST8-09"
/clone_1lb="Goat mammary gland"
/note="Mammary gland of Capra hircus (goat) "

BASE COUNT 161 a 95 c 94 g 159 t 9 others
ORIGIN

Query Match 7.5%; Score 18; DB 26; Length 518;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AATGCTGATGTTCTGTT 20
|||||
Db 61 AATGCTGATGTTCTGTT 44

RESULT 9
LOCUS AI022693 449 bp mRNA EST 28-AUG-1998
DEFINITION ow67609.x1 Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens CDNA
ACCESSION AI022693
NID 93237934
VERSION AI022693.1 GI:3237934
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 449)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR Gene Index
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2284689.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 2338 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 357.
location/Qualifiers
1. 449
/organism="Homo sapiens"

sequence.
ACCESSION 271911
NID 91292836
VERSION 271911.1 GI:1292836
KEYWORDS EST.
SOURCE goat.
ORGANISM Capra hircus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Caprinae; Capra.
REFERENCE 1 (bases 1 to 518)
AUTHORS Le Provost, F., Lepoint, A. and Martin, P.
TITLE A survey of the goat genome transcribed in the lactating mammary
gland
JOURNAL Mamm. Genome 7 (9), 657-666 (1996)
MEDLINE 96359152
COMMENT On Apr 14, 1993 this sequence version replaced gi:785308.

Contact: Le Provost, F.
Laboratoire de Genetique Biochimique
INRA-CRU
78352 Jouy-en-Josas Cedex, FRANCE.

FEATURES
source
1. 518
/organism="Capra hircus"
/db_xref="taxon:9925"
/clone_lib="EST8-09"
/clone_1lb="Goat mammary gland"
/note="Mammary gland of Capra hircus (goat) "

BASE COUNT 161 a 95 c 94 g 159 t 9 others
ORIGIN

Query Match 7.5%; Score 18; DB 41; Length 449;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 ATGGAATATCTACTTAT 234
|||||
Db 96 ATGGAATATCTACTTAT 79

RESULT 10
LOCUS T62858 572 bp mRNA EST 16-FEB-1995
DEFINITION yc03508.s1 Stratiagene lung (#937210) Homo sapiens CDNA clone
IMAGE:79575 3' similar to gb|L13712|HUMSCHLUR Human schRNA molecule,
transcribed from Alu (rRNA); gb:S41458 ROD CGMP-SPECIFIC
3',5'-CYCLIC PHOSPHODIESTERASE BETA-SUBUNIT (HUMAN);, mRNA
sequence.
ACCESSION T62858
NID 9666515
VERSION T62858.1 GI:666515
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 572)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, M.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, K., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevisan, E., Underwood, K., Woldmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
High quality sequence stops: 409
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Possible reversed clone: polyT not found
Seq primer: -21m13
High quality sequence stop: 409.
location/Qualifiers
1. 572
/organism="Homo sapiens"

sequence.
ACCESSION 271911
NID 91292836
VERSION 271911.1 GI:1292836
KEYWORDS EST.
SOURCE goat.
ORGANISM Capra hircus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Caprinae; Capra.
REFERENCE 1 (bases 1 to 518)
AUTHORS Le Provost, F., Lepoint, A. and Martin, P.
TITLE A survey of the goat genome transcribed in the lactating mammary
gland
JOURNAL Mamm. Genome 7 (9), 657-666 (1996)
MEDLINE 96359152
COMMENT On Apr 14, 1993 this sequence version replaced gi:785308.

Contact: Le Provost, F.
Laboratoire de Genetique Biochimique
INRA-CRU
78352 Jouy-en-Josas Cedex, FRANCE.

FEATURES
source
1. 518
/organism="Capra hircus"
/db_xref="taxon:9925"
/clone_lib="EST8-09"
/clone_1lb="Goat mammary gland"
/note="Mammary gland of Capra hircus (goat) "

BASE COUNT 161 a 95 c 94 g 159 t 9 others
ORIGIN

Query Match 7.5%; Score 18; DB 41; Length 449;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 ATGGAATATCTACTTAT 234
|||||
Db 96 ATGGAATATCTACTTAT 79

RESULT 10
LOCUS T62858 572 bp mRNA EST 16-FEB-1995
DEFINITION yc03508.s1 Stratiagene lung (#937210) Homo sapiens CDNA clone
IMAGE:79575 3' similar to gb|L13712|HUMSCHLUR Human schRNA molecule,
transcribed from Alu (rRNA); gb:S41458 ROD CGMP-SPECIFIC
3',5'-CYCLIC PHOSPHODIESTERASE BETA-SUBUNIT (HUMAN);, mRNA
sequence.
ACCESSION T62858
NID 9666515
VERSION T62858.1 GI:666515
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 572)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, M.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, K., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevisan, E., Underwood, K., Woldmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
High quality sequence stops: 409
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Possible reversed clone: polyT not found
Seq primer: -21m13
High quality sequence stop: 409.
location/Qualifiers
1. 572
/organism="Homo sapiens"

```

```

/organism="Homo sapiens"
/db_xref="GDB:483192"
/db_xref="taxon:9606"
/clone="IMAGE:79575"
/clone.lib="Stratagene lung (#937210)"
/sex="male"
/dev_stage="72 years"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. normal lung. Average insert size: 1.0 kb;
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGACGAC
3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'"

BASE COUNT      141 a      121 c      168 g      134 t      8 others
ORIGIN

Query Match      7.1%; Score 17; DB 21; Length 572;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      69 GGCACACTGTGAGAGAA 85
        |||||||
DB      218 GGCACACTGTGAGAGAA 234

RESULT 11
LOCUS      R42419      438 bp      mRNA      EST      22-MAY-1995
DEFINITION yf92e04.s1 Soares infant brain INTB Homo sapiens cDNA clone
IMAGE:29962 3' similar to gb:XS4156.fna1 CELLULAR TUMOR ANTIGEN P53
(HUMAN); contains Alu repetitive element; mRNA sequence.
ACCESSION      R42419
NID      9817185
VERSION      R42419.1 GI:817185
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 438)
AUTHORS      Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marrer,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
On May 8, 1995 this sequence version replaced gi:800643.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1628
High quality sequence stops: 344 Source: IMAGE Consortium, LLNT
This clone is available royalty-free through LLNT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1628 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 344.
Location/Qualifiers
1. .438
/organism="Homo sapiens"
/db_xref="GDB:402309"
/db_xref="taxon:9606"
/clone="IMAGE:29962"
/clone.lib="Soares infant brain INTB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"

```

```

/note="Organ: whole brain; Vector: latmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - Oligo(dt) primer 15'
ACTGAGAGAAATTCGCGCCGACAGATTTTCTTTTCTTTTCTTTT 3')
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT      122 a      91 c      88 g      134 t      3 others
ORIGIN

Query Match      7.1%; Score 17; DB 22; Length 438;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      125 AAAAAGTTAGCTGAAT 141
        |||||||
DB      108 AAAAAGTTAGCTGAAT 92

RESULT 12
LOCUS      AA112392      388 bp      mRNA      EST      08-NOV-1996
DEFINITION zn68g12.s1 Stratagene Hela cell s3 937216 Homo sapiens cDNA clone
IMAGE:563398 3' similar to contains Alu repetitive element; mRNA
sequence.
ACCESSION      AA112392
NID      91664802
VERSION      AA112392.1 GI:1664802
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 388)
AUTHORS      Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiappelli,B.,
Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Woldmann,P., Waterston,R., Wilson,R.
and Marrer,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham.
Location/Qualifiers
1. .388
/organism="Homo sapiens"
/db_xref="GDB:4593210"
/db_xref="taxon:9606"
/clone="21"
/clone.lib="Stratagene Hela cell s3 937216"
/sex="female"
/dev_stage="Hela S3 cell line"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcORI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. Hela S3
epitheloid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP XR

```





```

AF124520
ID AF124520 standard; RNA; EST; 361 BP.
XX
AC AF124520;
XX
SV AF124520.1
XX
NI 94433653
XX
DT 23-MAR-1999 (Rel. 59, Created)
DT 23-MAR-1999 (Rel. 59, Last updated, Version 1)
XX
DE Homo sapiens clone b5b6, mRNA sequence.
XX
EST.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC Primates; Catarrhini; Hominoidea; Homo.
XX
[1]
RN 1-361
RP Eickhoff B., van der Bosch J.;
RA "Identity of clones obtained from subtractive hybridization in human
RT lung adenocarcinoma cells";
RL Unpublished.
XX
[2]
RN 1-361
RP Eickhoff B., van der Bosch J.;
RA
RT Submitted (29-JAN-1999) to the EMBL/GenBank/DBJ databases.
RL Experimental Immunopharmacology, Research Center Borstel, Parkallee 22,
RL Borstel 23845, Germany
XX
XX
Key Location/Qualifiers
FH
FH source 1. 361
FT /db_xref="taxon:9606"
FT /note="similar to phospholipase C"
FT /organism="Homo sapiens"
FT /clone="b5b6"
FT /tissue_type="lung adenocarcinoma"
FT /cell_line="LR23"
FT /dev_stage="adult"
XX
SQ Sequence 361 BP; 97 A; 76 C; 112 G; 76 T; 0 other;

```

```

Query Match 6.7%; Score 16; DB 5; Length 361;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 77 TCGAGAGAGGCTGAA 92
DB 11 TCGAGAGAGGCTGAA 26

```

Search completed: August 6, 1999, 09:22:28  
Job time: 6219 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:09:58 ; Search time 3264.21 Seconds

(without alignments)  
265,983 Million cell updates/sec

Title: US-09-049-696-2

Perfect score: 273  
Sequence: 1 GTTGCATGACCCCATGT.....AGATGGCACTGTGGCGAG 273

Scoring table: IDENTITY\_NDC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBml:\*  
1: gb\_bal:\*  
2: gb\_bal:\*  
3: gb\_bal:\*  
4: gb\_bal:\*  
5: gb\_bal:\*  
6: gb\_bal:\*  
7: gb\_bal:\*  
8: gb\_bal:\*  
9: gb\_bal:\*  
10: gb\_bal:\*  
11: gb\_bal:\*  
12: gb\_bal:\*  
13: gb\_bal:\*  
14: gb\_bal:\*  
15: gb\_bal:\*  
16: gb\_bal:\*  
17: gb\_bal:\*  
18: gb\_bal:\*  
19: gb\_bal:\*  
20: gb\_bal:\*  
21: gb\_bal:\*  
22: gb\_bal:\*  
23: gb\_bal:\*  
24: gb\_bal:\*  
25: gb\_bal:\*  
26: gb\_bal:\*  
27: gb\_bal:\*  
28: gb\_bal:\*  
29: gb\_bal:\*  
30: gb\_bal:\*  
31: gb\_bal:\*  
32: gb\_bal:\*  
33: gb\_bal:\*  
34: gb\_bal:\*  
35: gb\_bal:\*  
36: gb\_bal:\*  
37: gb\_bal:\*  
38: gb\_bal:\*  
39: gb\_bal:\*  
40: gb\_bal:\*  
41: gb\_bal:\*  
42: gb\_bal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	267.4	97.9	3311	11	AF039400 Homo sapi

RESULT	1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS	AF039400					
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hclcal) mRNA, complete cds.					
ACCESSION	AF039400					
NID	94009457					
VERSION	AF039400.1					
KEYWORDS	GI:4009457					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.					
AUTHORS	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.					
TITLE	Genomic cloning, molecular characterization, and functional analysis of human hclcal, the first human member of the family of Ca <sup>2+</sup> -activated Cl <sup>-</sup> channel proteins					
JOURNAL	Genomics 54 (2), 200-214 (1998)					
MEDLINE	99047526					
REFERENCE	2 (bases 1 to 3311)					
AUTHORS	Gruber,A.D., Elble,R. and Pauli,B.U.					
TITLE	Direct Submission					

#### ALIGNMENTS

2	267.4	97.9	2826	42	AF127036	AF127036 Homo sapi
3	182.2	66.7	2937	12 <th>AB017156</th> <th>AB017156 Mus muscu</th>	AB017156	AB017156 Mus muscu
4	143	52.4	35278	11 <th>AF039401</th> <th>AF039401 Homo sapi</th>	AF039401	AF039401 Homo sapi
5	124.6	45.6	3317	3 <th>AF001261</th> <th>AF001261 Bos tauri</th>	AF001261	AF001261 Bos tauri
6	124.6	45.6	3288	3 <th>AF001262</th> <th>AF001262 Bos tauri</th>	AF001262	AF001262 Bos tauri
7	124.6	45.6	2820	3 <th>AF001263</th> <th>AF001263 Bos tauri</th>	AF001263	AF001263 Bos tauri
8	124.6	45.6	1265	3 <th>AF001264</th> <th>AF001264 Bos tauri</th>	AF001264	AF001264 Bos tauri
9	123	45.1	2984	3 <th>BT036445</th> <th>BT036445 Bos tauri</th>	BT036445	BT036445 Bos tauri
10	119.8	43.9	3415	42 <th>AF043976</th> <th>AF043976 Homo sapi</th>	AF043976	AF043976 Homo sapi
11	113.4	41.5	3022	12 <th>AF047838</th> <th>AF047838 Mus muscu</th>	AF047838	AF047838 Mus muscu
12	113.4	41.5	3471	12 <th>AF052746</th> <th>AF052746 Mus muscu</th>	AF052746	AF052746 Mus muscu
13	73.6	27.0	3604	9 <th>AB026833</th> <th>AB026833 Homo sapi</th>	AB026833	AB026833 Homo sapi
14	37.4	13.7	60791	11 <th>AC004749</th> <th>AC004749 Homo sapi</th>	AC004749	AC004749 Homo sapi
15	37	13.6	21180	4 <th>GDCOL6A1A</th> <th>X57998 G.domestica</th>	GDCOL6A1A	X57998 G.domestica
16	37	13.6	13213	7 <th>SCSOF</th> <th>X57950 S.cerevisia</th>	SCSOF	X57950 S.cerevisia
17	37	13.6	1590	7 <th>SCYKL224C</th> <th>228324 S.cerevisia</th>	SCYKL224C	228324 S.cerevisia
18	37	13.6	1810	7 <th>SCYKL225W</th> <th>228325 S.cerevisia</th>	SCYKL225W	228325 S.cerevisia
19	37	13.6	29634	8 <th>YSC19122</th> <th>U22383 Saccharomyc</th>	YSC19122	U22383 Saccharomyc
20	37	13.6	331326	34 <th>AC006802</th> <th>AC006802 Caenorhab</th>	AC006802	AC006802 Caenorhab
21	37	13.6	47147	37 <th>CELC16A11</th> <th>AF077536 Caenorhab</th>	CELC16A11	AF077536 Caenorhab
22	36	13.2	160754	11 <th>AC006213</th> <th>AC006213 Homo sapi</th>	AC006213	AC006213 Homo sapi
23	35.4	13.0	115641	8 <th>ATU95973</th> <th>U95973 Arabidopsi</th>	ATU95973	U95973 Arabidopsi
24	35.4	13.0	226505	34 <th>AC006899</th> <th>AC006899 Caenorhab</th>	AC006899	AC006899 Caenorhab
25	35.4	13.0	42100	36 <th>CEC52A11</th> <th>246792 Caenorhab</th>	CEC52A11	246792 Caenorhab
26	35.2	12.9	106329	8 <th>ATAC004411</th> <th>AC004411 Arabidops</th>	ATAC004411	AC004411 Arabidops
27	34.8	12.7	85096	35 <th>AC007323</th> <th>AC007323 Arabidops</th>	AC007323	AC007323 Arabidops
28	34.6	12.6	177503	11 <th>AC005297</th> <th>AC005297 Homo sapi</th>	AC005297	AC005297 Homo sapi
29	34.4	12.5	1203	7 <th>CAARS5</th> <th>X16634 Candida alb</th>	CAARS5	X16634 Candida alb
30	34.2	12.5	150668	34 <th>HS46038</th> <th>AL031662 Homo sapi</th>	HS46038	AL031662 Homo sapi
31	34.2	12.5	37845	36 <th>CEP01D4</th> <th>281054 Caenorhabd</th>	CEP01D4	281054 Caenorhabd
32	34	12.5	312430	1 <th>RPX02</th> <th>AJ235271 Rickettsi</th>	RPX02	AJ235271 Rickettsi
33	34	12.5	135815	10 <th>HS43408</th> <th>AL033504 Human DNA</th>	HS43408	AL033504 Human DNA
34	34	12.5	127824	11 <th>AC004822</th> <th>AC004822 Homo sapi</th>	AC004822	AC004822 Homo sapi
35	34	12.5	117864	34 <th>CEY3864</th> <th>299713 Caenorhabd</th>	CEY3864	299713 Caenorhabd
36	33.8	12.4	71618	7 <th>AB012240</th> <th>AB012240 Arabidops</th>	AB012240	AB012240 Arabidops
37	33.8	12.4	5235	7 <th>SCSTISG</th> <th>X74113 S.cerevisia</th>	SCSTISG	X74113 S.cerevisia
38	33.8	12.4	6440	7 <th>SCYOR153W</th> <th>275061 S.cerevisia</th>	SCYOR153W	275061 S.cerevisia
39	33.8	12.4	5240	7 <th>YSC1DRI</th> <th>D26548 Saccharomyc</th>	YSC1DRI	D26548 Saccharomyc
40	33.8	12.4	32396	8 <th>SC055020</th> <th>U55020 Saccharomyc</th>	SC055020	U55020 Saccharomyc
41	33.8	12.4	6552	8 <th>YSCPDR5X</th> <th>L19922 Saccharomyc</th>	YSCPDR5X	L19922 Saccharomyc
42	33.8	12.4	1865	10 <th>HSM800274</th> <th>AL049932 Homo sapi</th>	HSM800274	AL049932 Homo sapi
43	33.6	12.3	262336	34 <th>AC006786</th> <th>AC006786 Caenorhab</th>	AC006786	AC006786 Caenorhab
44	33.6	12.3	17564	7 <th>SCELM1A4</th> <th>X71821 S.cerevisia</th>	SCELM1A4	X71821 S.cerevisia
45	33.6	12.3	1804	7 <th>SCYKL040C</th> <th>228040 S.cerevisia</th>	SCYKL040C	228040 S.cerevisia

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES  
Source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1..3311  
/gene="hCLCA1"  
352..3096  
/note="transmembrane glycoprotein"  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPRKSSVFLLIHLLEGALSNSLIQLNNNGYEGIVVAIDPNP  
EDETLIQIKDMVTQASLYLEATGKRFYFNKVALIPETWTKADYVPRLEYKNA  
DLVAESTIPGNDEPYTEONGNCGEKERHLPDFIAGKRLAEYPGKAFVEMAH  
LRMGVFEYNDDEKYLNSGRIOAVRCSAGITGVYKCGGSCYTRCFNKVTGL  
YKGEFEVLSROTEKASIMFAOHVDSIVECTEONHNKKAAPNOKNCKLRSTWEVI  
RDSDEKFTPTMTOPNPTEFSLOIORIVCLVDKSGSMATGNRLRLNOGOLF  
LQYBELSGWGMVTFDSAAVOSLIIQINSGSDRLAKRLPAASGCTICSGLRSA  
FTYIRKRYPTDGSSEIVLLTDGENDTISGCENKQSAIHTVALGSAOELBELSK  
MTGGLQTYASDOVONNGLIDAFGLSNGNVAVSORSIOLSKGLTLQNSOMNVTIV  
DSYVGKDTLELTWTTPQIILMDPSGOKGQGVNDKTMAYLIQIPGAKVGTWKY  
SLOASQTLTLVTSRASNAITLPITVTSKTKSPSPPLVYANIRGASPIILRA  
SVTALISVNGKTYLLELDNGAGADATKDGVSRYFTYDNGRSYVKRALGVN  
AARRVYPOOGSALYIPGIMTENDIOMNPRPEINKDVCHKQVCFSTSGSFVSA  
DVNAPIPDLPPGOITDLKAEIHGSLINFTAPGDDYDHGAHKITLIRISYILD  
LRDKNESLQVNTTALLIPKRNSEVFLFEPNTFFENGDTLFLAIOAVDVLDKSLI  
SNIRVSLFIPQTPPETPSPDETSAPCPNINHINSTITPGIHLIKIMKMGIELDLSIA

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 97.9%; Score 267.4; DB 11; Length 3311;  
Best Local Similarity 98.2%; Pred. No. 8.3e-61;  
Matches 268; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GTTGAATCGACCCCAATGTGCCAGAAAGTGAACACTATTCACAAATAAAGACATG 60  
DB 460 GTTGAATCGACCCCAATGTGCCAGAAAGTGAACACTATTCACAAATAAAGACATG 519  
OY 61 GTGACCCAGGATCTCTGATCTGTTGAAGCTACAGAAAGCAATTTATTCAAAAAT 120  
DB 520 GTGACCCAGGATCTCTGATCTGTTGAAGCTACAGAAAGCAATTTATTCAAAAAT 579  
OY 121 GTTGCATTTGATTCCTGAACAATGAGAAAGAGTGAATGTGAGACCAAACTT 180  
DB 580 GTTGCATTTGATTCCTGAACAATGAGAAAGAGTGAATGTGAGACCAAACTT 639  
OY 181 GAGACCTACAAAAATGCTGATGTCTGCTGAGTCTAATCCCTCCAGGNAATGATGA 240  
DB 640 GAGACCTACAAAAATGCTGATGTCTGCTGAGTCTAATCCCTCCAGGNAATGATGA 699  
OY 241 CCTACACTGNGCAGATGGGCAACTGTGGCGAG 273  
DB 700 CCTACACTGAGCAGATGGGCAACTGTGGAGAG 732

RESULT 2  
AF127036 2826 bp mRNA PRI 16-Apr-1999  
LOCUS AF127036  
DEFINITION Homo sapiens caldium-activated chloride channel protein 1 (cacc1)  
ACCESSION AF127036  
NID 94585468  
VERSION AF157036.1 GI:4585468

KEYWORDS

SOURCE

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
AUTHORS Agnel M., Vermet T., and Culouscou J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel M. and Culouscou J.-M.  
TITLE Direct Submission

JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Rueil-Malmaison 92500, France

FEATURES  
source

gene  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1..2826  
/gene="cacc1"  
5..2749  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="A425487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPRKSSVFLLIHLLEGALSNSLIQLNNNGYEGIVVAIDPNP  
EDETLIQIKDMVTQASLYLEATGKRFYFNKVALIPETWTKADYVPRLEYKNA  
DLVAESTIPGNDEPYTEONGNCGEKERHLPDFIAGKRLAEYPGKAFVEMAH  
LRMGVFEYNDDEKYLNSGRIOAVRCSAGITGVYKCGGSCYTRCFNKVTGL  
YKGEFEVLSROTEKASIMFAOHVDSIVECTEONHNKKAAPNOKNCKLRSTWEVI  
RDSDEKFTPTMTOPNPTEFSLOIORIVCLVDKSGSMATGNRLRLNOGOLF  
LQYBELSGWGMVTFDSAAVOSLIIQINSGSDRLAKRLPAASGCTICSGLRSA  
FTYIRKRYPTDGSSEIVLLTDGENDTISGCENKQSAIHTVALGSAOELBELSK  
MTGGLQTYASDOVONNGLIDAFGLSNGNVAVSORSIOLSKGLTLQNSOMNVTIV  
DSYVGKDTLELTWTTPQIILMDPSGOKGQGVNDKTMAYLIQIPGAKVGTWKY  
SLOASQTLTLVTSRASNAITLPITVTSKTKSPSPPLVYANIRGASPIILRA  
SVTALISVNGKTYLLELDNGAGADATKDGVSRYFTYDNGRSYVKRALGVN  
AARRVYPOOGSALYIPGIMTENDIOMNPRPEINKDVCHKQVCFSTSGSFVSA  
DVNAPIPDLPPGOITDLKAEIHGSLINFTAPGDDYDHGAHKITLIRISYILD  
LRDKNESLQVNTTALLIPKRNSEVFLFEPNTFFENGDTLFLAIOAVDVLDKSLI  
SNIRVSLFIPQTPPETPSPDETSAPCPNINHINSTITPGIHLIKIMKMGIELDLSIA

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 97.9%; Score 267.4; DB 42; Length 2826;  
Best Local Similarity 98.2%; Pred. No. 8.3e-61;  
Matches 268; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GTTGAATCGACCCCAATGTGCCAGAAAGTGAACACTATTCACAAATAAAGACATG 60  
DB 113 GTTGAATCGACCCCAATGTGCCAGAAAGTGAACACTATTCACAAATAAAGACATG 172  
OY 61 GTGACCCAGGATCTCTGATCTGTTGAAGCTACAGAAAGCAATTTATTCAAAAAT 120  
DB 173 GTGACCCAGGATCTCTGATCTGTTGAAGCTACAGAAAGCAATTTATTCAAAAAT 232  
OY 121 GTTGCATTTGATTCCTGAACAATGAGAAAGAGTGAATGTGAGACCAAACTT 180  
DB 233 GTTGCATTTGATTCCTGAACAATGAGAAAGAGTGAATGTGAGACCAAACTT 292  
OY 181 GAGACCTACAAAAATGCTGATGTCTGCTGAGTCTAATCCCTCCAGGNAATGATGA 240  
DB 293 GAGACCTACAAAAATGCTGATGTCTGCTGAGTCTAATCCCTCCAGGNAATGATGA 352  
OY 241 CCTACACTGNGCAGATGGGCAACTGTGGCGAG 273  
|||||

```

Db      353  CCTACACTGAGCAGATGGCGAACAATGTTGAGAG 385

RESULT  3
LOCUS   AB017156
DEFINITION Mus musculus gob-5 mRNA, complete cds.
ACCESSION AB017156
NID      93721911
VERSION  AB017156.1  GI:3721911
KEYWORDS GOB-5.
SOURCE   Mus musculus adult intestine goblet cell cDNA to mRNA.
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Mammalia;
          Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.
TITLE    Cloning and identification of the gene gob-5, which is expressed in
          intestinal goblet cells in mice
JOURNAL  Biochem. Res. Commun. 255 (2), 347-351 (1999)
MEDLINE  99160866
REFERENCE 2 (bases 1 to 2937)
AUTHORS   Komiyama, T., Tanigawa, Y. and Hirohashi, S.
TITLE     Direct Submission
JOURNAL   Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru
          Komiyama, ERATO, JST, Hirohashi Cell Configuration Project. 5-9-9,
          Tokodai, Tsukuba, Ibaraki 300-2655, Japan
          (E-mail: tkomihccp.jst.go.jp, Tel:81-298-47-7563,
          Fax:81-298-47-5226)

FEATURES
source   location/Qualifiers
         1..2937
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /cell_type="goblet cell"
            /dev_stage="adult"
            /tissue_type="intestine"
            15..2756
            /gene="gob-5"
            15..2756
            /gene="gob-5"
            /codon_start=1
            /protein_id="BAA33743.1"
            /db_xref="PID:d1034712"
            /db_xref="PID:g3721912"
            /db_xref="GI:3721912"
            /translation="MESLKSPEVFLIIHLLEGVISELIQNLNNNGTCIYAIADHYD
            /EDLELQIHKDMVTQASPVYLEATGCKRFEFNVAIILIPESKAPETRPRLTEFNAA
            DVALSTSPLENDDEPYEHLIGACENKGRIRHLPDPLAGKLTQYGDQDFVEMAH
            FRMVFENYNNDEKFLYLSKGPQAVRCSAATGKNQVRCOGSGCINRCACVDIRVYG
            LYKNCVFEPPPHONEKASITFENQINNSVVEFCENKNDAPDOROKRSTSEV
            IOESDEPKOTTPMTPAOSAPAPPEFSELTQGRVCLVILDKSGMLNDDRLNHNNAARLE
            LLQJVEGDSWGWGATPDSAAVYQSELIQKNSGADRLILKHLPIVYSAGGTSICGELTPT
            AFYIIRKRYPLPTDGEIVLLTDEONTISSCPDLVKGSAIHHVALGPAAKLEQLS
            KMGQDTSDDQVONNGLVDAFALNSGNAIQQHSIOLESRCVNLQNNMNGNSVYI
            VDSVGKDTLEFLITWTHHPPIFLIMDSGVQNGFIIDITTKVYLVQPGAKGFYWK
            YSIOASSOTLITLVTSRAASATLPIPIYTPVKNNTGKPEPPIVVASIRGASPIILR
            ASYALILSNVNGKVTIVLELDNGAGADATKRDGYSRFPFADNGKSYKTYALGCV
            TSDQORAAPKRNKMATIDGWLDEGEVRNRPFRPTSTVQDQLCFRTSSGGSFVATIN
            VPAAAPLIPDLPPCQIIDLKASIQGQNLVNLTPWAPDDYDHGASNYILRMSTSIYDV
            LRDAHPTSLQVNTITGLIPKEASSESEIEFEELGAGTIFNGTDIFAIQAVKSNLKSSEI
            SNIRASVFIPAOEPPPEIDSTPCPDPISTINSITPIGHVILKIMKMLGEMQVTLGLH"

BASE COUNT      860 a      718 c      693 g      666 t

ORIGIN

```

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	JOURNAL	FEATURES	source
62	GCACCCAGGACATCTCTGATCTGTTTGAAGTACAGGAAAGCGATTTTATTTCAAATG 121					human.										
184	TGACTCAGGCTCTTCACATACCTGTTTAAAGCTACAGGAAAAAGATTCTTCAAATG 243					human.										
122	TTGGCATTTTGTATCTCTGAAACATGAGAAGCAAAAGTGTACTATGTGAGACCAAACTTG 181					human.										
244	TTGGCATTTTGTATCTCCGAGAGCTGGAAGGCAAAAGCCTGATATATACGAGCCAAACTTG 303					human.										
182	AGACCTACAAAAATGCTGATGTTGTTGCTGCTGAGTATTCCTCCAGGAAATGATGAC 241					human.										
304	AAACCTTCAAAAAAGCGTATGTCCTTTATACAAACACACACCCCTCTAGGCAATGATGAGC 363					human.										
242	CCTACACTGNGCAGATGGGCAACTGTGGCG 272					human.										
364	CCTACACCGAACAATATAGAGCATGTGAGGA 394					human.										
AF039401	AF039401 35278 bp DNA	PRI	15-DEC-1998			human.										
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene, complete cds.					human.										
ACCESSION	AF039401					human.										
NID	94009459					human.										
VERSION	AF039401.1	GI:4009459				human.										
KEYWORDS						human.										
SOURCE						human.										
ORGANISM	Homo sapiens					human.										
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					human.										
AUTHORS	1 (bases 1 to 35278)					human.										
TITLE	Gruber,A.D., Elble,R.C., Ji,H.L., Schneur,K.D., Fuller,C.M. and Pauli,B.U.					human.										
JOURNAL	Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins					human.										
MEDLINE	Genomics 54 (2), 200-214 (1998)					human.										
REFERENCE	2 (bases 1 to 35278)					human.										
AUTHORS	Gruber,A.D., Elble,R. and Pauli,B.U.					human.										
TITLE	Direct Submission					human.										
JOURNAL	Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA					human.										
FEATURES	Location/Qualifiers					human.										
source	1. 35278					human.										
	/organism="Homo sapiens"					human.										
	/db_xref="taxon:9606"					human.										
	/chromosome="1"					human.										
	/map="1p22-p31"					human.										
	1. 1617					human.										
	/gene="hCLCA1"					human.										
	1. 33522					human.										
	/gene="hCLCA1"															

24532..24628.26644..36859.27447..27708.28766..28936,  
31840..32079.32919..33310)  
/gene="hclcal1"  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPEKSSVFLLIHLLEGALSNLIQINNGYEGIVVADPNVP  
EDELIIQIKDMVTOASLYLEFATGKRRYFKNVAILPETWKTADYRPLLEYKNA  
DLVAESIPYNDDEPTEOMGNGCEGRILITPDLAGKLAEGCGKAFVEMH  
LRMGVPEYNNDEKFEYLSNGRIQAVRGAGITGNVYKCGGSCYTRCFENVTGL  
YEKCEFTLOSROTEKASIMFAOHVDSVECTEONHKEA.PNKONOKN.RSTWEYI  
RDSDEKFTTPMTTOPNPPTSLSQIGRTICVLDSGSMATNRRLRLNQAQDEL  
LQVLELSWGMVTFDSAAHVQSELIQINSQSDDTLAKRLPAASGGTSLCSGLRSA  
FTVIRKRYPTDGEISVILTDGENDTISCFNEVQSGAIITHTVALGSAQDELSEK  
MTGLOIYASDOVONNGILDAFGALSGNCAVSORIOLESKGLTNSQMANGTIVY  
DSYVGKDTLFLITWTOPPOLIMDPDSGQGVKNTKMAVIOIPIGKVGTMKY  
SLQASQDTLITVSRAKNATLPITYTSKNTKTSKPSPLIYANTROGASPIIISA  
SVTALIESVNGRTVLELLDNGAGADTKDGVSRFTITDNGRISVAKALGVNA  
AARRVIFQOQSGALYIPGMIENDEIQWNPPEINKDVQKQVCFSEKSGFVAS  
DVPNAPIDLPFGQITDLNAEIHGSLINTWTAPGDYDGHAKXIIIRISTIID  
LRDKENISLOVNTALIPKEANSEVFLFKEPENTITFEGDTLFAIOAVDVDSKSEI  
SNIRVSLFIPQTPPEPSPDETSAPCPNINHINSTITGHIILKIMMWIGELSLISA  
"  
6667..6807  
/gene="hclcal1"  
/number=3  
6979..7126  
/gene="hclcal1"  
/number=4  
9696..9801  
/gene="hclcal1"  
/number=5  
15454..15631  
/gene="hclcal1"  
/number=6  
18591..18809  
/gene="hclcal1"  
/number=7  
19783..20010  
/gene="hclcal1"  
/number=8  
22253..22427  
/gene="hclcal1"  
/number=9  
24522..24628  
/gene="hclcal1"  
/number=10  
26644..26859  
/gene="hclcal1"  
/number=11  
27447..27708  
/gene="hclcal1"  
/number=12  
28766..28936  
/gene="hclcal1"  
/number=13  
31840..32079  
/gene="hclcal1"  
/number=14  
32919..33522  
/gene="hclcal1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN

Query Match 52.4%; Score 143; DB 11; Length 35278;  
Best Local Similarity 99.3%; Pred. No. 3.6e-28;  
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 53 AGGACATGGTGCCAGGACATCTCTGTATCTGTGAAGCTACAGGAAGCATTTTAT 112

|||||  
Db 6665 AGGACATGGTGCCAGGACATCTCTGTATCTGTGAAGCTACAGGAAGCATTTTAT 6724  
QY 113 TCAAAATGTTGCCATTTTGAATTCCTGAACATGGAAGACAAAGNTGACTATGTGAGAC 172  
Db 6725 TCAAAATGTTGCCATTTTGAATTCCTGAACATGGAAGACAAAGNTGACTATGTGAGAC 6784  
QY 173 CAAACTGACACTACAAAATG 196  
Db 6785 CAAACTGACACTACAAAATG 6808  
RESULT 5  
AF001261  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNML  
FEATURES  
CDS  
BASE COUNT 1084 a 676 c 634 g 923 t  
ORIGIN

AF001261 3317 bp mRNA NAM 19-NOV-1997  
Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
complete cds.  
AF001261 67623762  
AF001261.1 GI:2623762  
Bos taurus.  
Bos taurus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
1 (bases 1 to 3317)  
Eidley, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
Cloning and characterization of Lu-ECAM-1 suggest it is an  
endothelial chloride channel  
Unpublished  
2 (bases 1 to 3317)  
Eidley, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
Direct Submission  
Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA  
Location/Qualifiers  
1..3317  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="1"  
/tissue\_type="lung"  
63..2780  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86529.1"  
/db\_xref="PID:92623763"  
/db\_xref="GI:2623763"  
/translation="MVLCLNVILFLHLHPGKSSNVNLINNGYDGVIAINPSVE  
DEKLENIKEMVTEASTYLPHATKRYFENVAILIPMTKSKSEFIPQESYDQAD  
VIVANPLYKGGDDPYTLOYRGGEKGYIHFTNLTNNHGYISGRYFVEMHL  
RMGIDPEYNDQPEFYISRKRTIATRESTITGILINVEKCGGSGCTSLCRDSDQIG  
LYEAKCTFLPKSQTAKESTIMFMSLSVTEFCTEXTHTNTEAPRLQKMGKNGSTMDY  
INNSVDEQNTSPMTENMPPTHTFSLKSQRYVACVLDSGSMASADRLEPQNOAAE  
LYLIQVLEKSLGVMVTFDSVAIEIONLIRITPDNYOKITAKLPQVANGTISICRL  
KAGFOALIHSDQSTGSEIILTDGENDTISCFNEVQSGAIITHTVALGSAQDELSEK  
TLNMGYGVYFANKDITGLTNAFSRISRSQSGITQOALQESKALKITGRKRVNGV  
PVDSTVGNDFEYVYWTIQPEYVLDPPKRRKKTISFKEDKLNISARQIQIGIAET  
GWTYISLNNHASSQMLTVTTTRARSPITPVIAIAMSQHTAHPSPHIVIAQVSQ  
GVLPGVIGISVIALIETEDHQVTLLEMDNGAGDYKNDIYRSYFTDYDGNORYSLK  
VHAQARNNTARLNLROPNKVLVYGVYENGKIIINPREVVDLAKAKIEDFSRLT  
SGGSFTVSGAPPGNHSVPPSKITDLEAKFKEDYIOLSTAPGNVLDRKANSYIIR  
RISKSEMDROEDDNATIVNTSNLIPKASKEKPEKPEHFEVNGTKRYISVQALN  
EANLISEVSHIVQAIKFIPLPDSYHDLGKIKISEITLAIIGLMIFSIV"

Query Match 45.6%; Score 124.6; DB 3; Length 3317;



```

/db_xref="GI:2623767"
/translation="MVCILNVILFLTLHLPGKSSMVLINNGDGIYAIINPVE
DEKLEINIKEMVTEASTYLPHATRRVYRNISILPMWKSSEYFIRKQESYDQAD
VIVANPYLKTIGDDPYTLQYRGCEKGYIHETPNFLITNNPHIYSGRPFVEMAH
RMGIFDEYNVDOPYISRNKTEATRCSTHITGVKCPGSCITSLCRDSQTV
LYEAKCTFLPKSQTAKEIMFPLSLKSVTECTEKHTNTEAPNLQNKCKNGKSTWDV
IMNSVDONTSPMTENMPPTHTPRLSKRVCLVLDKSGSMSADRLFOVNOQAE
LYLQVLEKSGLVGMVETDVAEITONHLRTIDNDVOKITAKLPQVANGSTICRGL
KAGFOALIHDSQTSSEIILLTDGENEINSCPEYKRSGLIHTIATLGPAAKELE
PLSMVGTGIRFPFANKDITGLTNARSRSBSGISTOALOLESKALITGRKRYNGV
PVDSTVGNDIFFVVTWIIQPEIVLDKGGKRYTSPEKDKLIRSAQOIBIAT
GTWYSLVNLNHSQMLVTVTRASPTLPPVATAHMSQHTAHYSPMVAQVQ
GFLVPLGISVIAIETEDGHQVTLMDNGARDVNDGYSRYFTDYGNGYSLSK
VHAORNTARLNRPONKVLVYGVYENGKILNPPREPKDLAKAKIEDPSRLT
SGSPFTSGAPPNPBPVPPSKITDLEAFKEDYIQLSWTAVGNLIDKXKASGSP
MSRSHQVAKLEIQLQHQSRQ"

```

BASE COUNT 940 a 565 c 560 g 755 t  
ORIGIN

Query Match 45.6%; Score 124.6; DB 3; Length 2820;  
Best Local Similarity 65.7%; Pred. No. 2.6e-23;  
Matches 178; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```

OY 2 TTGCATCGACCCCAATGTGCCAGAGATGAACACTATTCACAAATAAAGACATGG 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 300 TTGCATTAACCCAGATGTGCCAGAGATGAACACTATTCACAAATAAAGAAATGG 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 62 TGACCCAGCATCTCTGTATCTGTTTGAAGCTACAGAAAGGATTTTATTCAAAATG 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 360 TAACTGAAGCTTCTACTTACCTGTTTATGACCAACAAAGAGAGTTTATTCAGAAATG 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 122 TTGCCATTTTGATCTGGAACATGGAAGCAAAAGSGTACTATGTGAGACCAAACTTG 181
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 420 TGAGCATTTTAAATTCATGACCTGGAATCAAAATCTGATGATCATACCAAAACAG 479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 182 AGACCTACAAAATGCTGATGTTCTGCTGAGTCTANTCTCCAGGNAATGATGAAAC 241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 480 AATCATATGACCGACAGATGTCATGATGCTATCCATTAATAATGAGATGATC 539
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 242 CCTACACTGNGCAGATGGGCAACTGTGGCGA 272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 540 CCTATACACTCAATATGGAAGGTGTGAGA 570
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 8  
AF001264 1265 bp mRNA MAM 19-NOV-1997  
LOCUS Bos taurus clone 4 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.

ACCESSION AF001264  
NID 92623768  
VERSION AF001264.1 GI:2623768  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.

REFERENCE 1 (bases 1 to 1265)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Chany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
endothelial chloride channel  
JOURNAL unpublished  
REFERENCE 2 (bases 1 to 1265)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Chany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA

FEATURES  
source 1..1265  
Location/Qualifiers

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue_type="lung"
/cell_type="aortic endothelial cells"
/clone="4"
109..1137
/note="endothelial adhesion molecule; chloride channel;  
secreted form"
/codon_start=1
/product="Lu-ECAM-1"
/protein_id="AAB8532.1"
/db_xref="PIR:Q2623769"
/db_xref="GI:2623769"
/translation="MVCILNVILFLTLHLPGKSSMVLINNGDGIYAIINPVE
DEKLEINIKEMVTEASTYLPHATRRVYRNISILPMWKSSEYFIRKQESYDQAD
VIVANPYLKTIGDDPYTLQYRGCEKGYIHETPNFLITNNPHIYSGRPFVEMAH
RMGIFDEYNVDOPYISRNKTEATRCSTHITGVKCPGSCITSLCRDSQTV
LYEAKCTFLPKSQTAKEIMFPLSLKSVTECTEKHTNTEAPNLQNKCKNGKSTWDV
IMNSVDONTSPMTENMPPTHTPRLSKRVCLVLDKSGSMSADRLFOVNOQAE
LYLQVLEKSGLVGMVETDVAEITONHLRTIDNDVOKITAKLPQVANGSTICRGL
KAGFOALIHDSQTSSEIILLTDGENEINSCPEYKRSGLIHTIATLGPAAKELE
PLSMVGTGIRFPFANKDITGLTNARSRSBSGISTOALOLESKALITGRKRYNGV
PVDSTVGNDIFFVVTWIIQPEIVLDKGGKRYTSPEKDKLIRSAQOIBIAT
GTWYSLVNLNHSQMLVTVTRASPTLPPVATAHMSQHTAHYSPMVAQVQ
GFLVPLGISVIAIETEDGHQVTLMDNGARDVNDGYSRYFTDYGNGYSLSK
VHAORNTARLNRPONKVLVYGVYENGKILNPPREPKDLAKAKIEDPSRLT
SGSPFTSGAPPNPBPVPPSKITDLEAFKEDYIQLSWTAVGNLIDKXKASGSP
MSRSHQVAKLEIQLQHQSRQ"
FKLIGNTI"

```

BASE COUNT 440 a 231 c 231 g 363 t  
ORIGIN

Query Match 45.6%; Score 124.6; DB 3; Length 1265;  
Best Local Similarity 65.7%; Pred. No. 2.6e-23;  
Matches 178; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```

OY 2 TTGCATCGACCCCAATGTGCCAGAGATGAACACTATTCACAAATAAAGACATGG 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 215 TTGCATTAACCCAGATGTGCCAGAGATGAACACTATTCACAAATAAAGAAATGG 274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 62 TGACCCAGCATCTCTGTATCTGTTTGAAGCTACAGAAAGGATTTTATTCAAAATG 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 275 TAACTGAAGCTTCTACTTACCTGTTTATGACCAACAAAGAGAGTTTATTCAGAAATG 334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 122 TTGCCATTTTGATCTGGAACATGGAAGCAAAAGSGTACTATGTGAGACCAAACTTG 181
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 335 TGAGCATTTTAAATTCATGACCTGGAATCAAAATCTGATGATCATACCAAAACAG 394
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 182 AGACCTACAAAATGCTGATGTTCTGCTGAGTCTANTCTCCAGGNAATGATGAAAC 241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 395 AATCATATGACCGACAGATGTCATGATGCTATCCATTAATAATGAGATGATC 454
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 242 CCTACACTGNGCAGATGGGCAACTGTGGCGA 272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 455 CCTATACACTCAATATGGAAGGTGTGAGA 485
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 9  
BT036445 2984 bp mRNA MAM 09-FEB-1996  
LOCUS Bos taurus calcium-activated chloride channel mRNA, complete cds.  
DEFINITION U36445  
ACCESSION U36445  
NID 91184065  
VERSION U36445.1 GI:1184065  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.

REFERENCE 1 (bases 1 to 2984)  
AUTHORS Cunningham, S.A., Awaide, M.S., Budien, J.K., Ismailov, I.I.,  
Arrate, M.P., Berdley, B.K., Benos, D.J. and Fuller, C.M.  
TITLE Cloning of an epithelial chloride channel from bovine trachea  
JOURNAL J. Biol. Chem. 270 (52), 31016-31026 (1995)  
REFERENCE 2 (bases 1 to 2984)  
AUTHORS Benos, D.J.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,  
University of Alabama at Birmingham, BHSB 706, Birmingham, AL

FEATURES  
source 1..2984  
Location/Qualifiers



FEATURES 35294, USA  
 source Location/Qualifiers  
 1. .2984  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /cell\_type="epithelial cells"  
 /tissue\_type="trachea"  
 19. .2730  
 /note="epithelial chloride channel"  
 /codon\_start=1  
 /product="calcium-activated chloride channel"  
 /protein\_id="AAC48511.1"  
 /db\_xref="PID:91184066"  
 /db\_xref="GI:1184066"  
 /translation="MPERLVYLFLTLHLPEKSSMYNLNNGIDGVIATINPSVE  
 DEKLIONIKEMVETASTYLPHATKRRVYFRNVSILIPMTWSKSEYLPKDESYDAE  
 VIVANPYLKHGDDPYTLOYRGCEKGOYIHTPNFLTNLPIYSGRAVEHWAHL  
 RMGIFDEYNQDPYISRRNTIETRCSTHITGVIVKCGSCITPRCRDSOTGL  
 YEACCTFIPEKSOTAKESIMWOSLHSTPECTEKTHTVVEAPNQLNKKMGSTMDVI  
 MNSDFONTSPTEENPPTQPTFFSLKSKORVCIYLDKSGSSEDLFRNQAEL  
 FLIOIIEGSLGVATFDSVAEIRNNLTKITDNYENITANLPOEANGTISICGLK  
 AGFOAIIOOOSTSGSEIILLDGEDNEHSCIEVKOSGVIIHTIALGSPSKALET  
 LSDMGHGFYAKNDINGLNFASISRSRSGSITQOTIOLESKALATEKKNVETVP  
 VDSITGNDPEFVYWTIKKPEILLQDPKGRKXTSDPEKEDKLINHSARLRIPGIEVG  
 TWTYSLLNNHASPOLLIVTVTRASPPTPTATAHMSONTAHPSPVIYAYVSOG  
 FLPIVGLIVTALIEDEGHQVLELDNMGAGADILPKNGITSRTITDYGNGRISLKV  
 HAEARNNTARLSLROPONKALYIPGIYENGKILNPPPEVADLAKAIEDFESLTS  
 GGSFVDSGAPPNHPSVLPNPKIIDLKAEKEDHQLQMTAPANVLDGKANSYIIRI  
 SKSFLLDKEDNATLVNTSLKPKKAGSDEFEKPPFPRIENGTNFIYAVOAINEA  
 NLTSEVSNIAQKRIIPMEDSDVPALGTRKISINAIIFALMILISIV"

BASE COUNT 1027 a 583 c 561 g 813 t  
 ORIGIN

Query Match 45.1%; Score 123; DB 3; Length 2984;  
 Best Local Similarity 65.3%; Pred. No. 6.7e-23;  
 Matches 177; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 2 TTGCATCGACCCCAATGTCGAGAGATGAACACTATTCACAAATPAAGACATGG 61  
 DB 125 TTGCATTAATCCAGTGTGACGAGAGATGAACACTATTCACAAATPAAGAAATGG 184  
 QY 62 TGACCCAGGACATCTGATCTGTTGAGCTACAGAAAGCATTTTTCACAAATG 121  
 DB 185 TAACGAGACCTTCTACTTACCTGTTTCATGCCAACAGAAAGATTATTTTCAGGAATG 244  
 QY 122 TTGCATTTGATTCCTGAGACATGGAAGAGAGTGTGAGACCAAAACTTG 181  
 DB 245 TAAGCATTTTAAATTCATAGAGAGTGAATCAAAATCTGATCTTAATGCCAAAATG 304  
 QY 182 AGACCTACAAAATGCTGATGTTGCTGCTGAGTCTAATCCTCAGGNAATGATGAAC 241  
 DB 305 AATCATATGACGAGAGATGATGCTAATCTTACCTAATAAATGAGATGATC 364  
 QY 242 CCTACACTGNGCAGATGGCAACTGTGGCGA 272  
 DB 365 CCTATACACTCAATATGAGAAATGTGGAGA 395

RESULT 10  
 AF043976 3415 bp mRNA PRI 07-APR-1999  
 LOCUS Homo sapiens CLCA homolog (hCLCA3) mRNA, complete cds.  
 DEFINITION AF043976  
 ACCESSION AF043976  
 NID 94572288  
 VERSION AF043976.1 GI:4572288  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 3415)  
 AUTHORS Gruber,A.D. and Pauli,B.U.

TITLE Molecular cloning and biochemical characterization of a truncated,  
 secreted member of the human family of Ca<sup>2+</sup>-activated Cl<sup>-</sup> channels  
 JOURNAL Biochim. Biophys. Acta 1444 (3), 418-423 (1999)  
 MEDLINE 99196715  
 REFERENCE 2 (bases 1 to 3415)  
 AUTHORS Gruber,A.D., Elble,R.C. and Pauli,B.U.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of  
 Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

FEATURES  
 source Location/Qualifiers  
 1. .3415  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1. .3415  
 /gene="hCLCA3"  
 /gene="hCLCA3"  
 19. .807  
 /codon\_start=1  
 /product="CLCA homolog"  
 /protein\_id="AAC23734.1"  
 /db\_xref="PID:94572289"  
 /db\_xref="GI:4572289"  
 /translation="MPESLKYLFLSLSPVLSKSLVTLNNGYDGIATINPSVE  
 DEKLIONIKEMVETASTYLPHATKRRVYFRNVSILIPMTWSKSEYLPKDESYDAE  
 VIVADLYIKYDDEPYTLOYRGCEKGOYIHTPNFLTNLPIYSGRAVEHWAHL  
 RMGIFDEYNQDPYISRRNTIETRCSTHITGVIVKCGSCITPRCRDSOTGL  
 YEACCTFIPEKSOTAKESIMWOSLHSTPECTEKTHTVVEAPNQLNKKMGSTMDVI  
 MNSDFONTSPTEENPPTQPTFFSLKSKORVCIYLDKSGSSEDLFRNQAEL  
 FLIOIIEGSLGVATFDSVAEIRNNLTKITDNYENITANLPOEANGTISICGLK  
 AGFOAIIOOOSTSGSEIILLDGEDNEHSCIEVKOSGVIIHTIALGSPSKALET  
 LSDMGHGFYAKNDINGLNFASISRSRSGSITQOTIOLESKALATEKKNVETVP  
 VDSITGNDPEFVYWTIKKPEILLQDPKGRKXTSDPEKEDKLINHSARLRIPGIEVG  
 TWTYSLLNNHASPOLLIVTVTRASPPTPTATAHMSONTAHPSPVIYAYVSOG  
 FLPIVGLIVTALIEDEGHQVLELDNMGAGADILPKNGITSRTITDYGNGRISLKV  
 HAEARNNTARLSLROPONKALYIPGIYENGKILNPPPEVADLAKAIEDFESLTS  
 GGSFVDSGAPPNHPSVLPNPKIIDLKAEKEDHQLQMTAPANVLDGKANSYIIRI  
 SKSFLLDKEDNATLVNTSLKPKKAGSDEFEKPPFPRIENGTNFIYAVOAINEA  
 NLTSEVSNIAQKRIIPMEDSDVPALGTRKISINAIIFALMILISIV"

BASE COUNT 1149 a 625 c 630 g 1011 t  
 ORIGIN

Query Match 43.9%; Score 119.8; DB 42; Length 3415;  
 Best Local Similarity 64.6%; Pred. No. 4.6e-22;  
 Matches 175; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 2 TTGCATCGACCCCAATGTCGAGAGATGAACACTATTCACAAATPAAGACATGG 61  
 DB 125 TTGCATTAATCCAGTGTGACGAGAGATGAACACTATTCACAAATPAAGAAATGG 184  
 QY 62 TGACCCAGGACATCTGATCTGTTGAGCTACAGAAAGCATTTTTCACAAATG 121  
 DB 185 TAACGAGACCTTCTACTTACCTGTTTCATGCCAACAGAAAGATTATTTTCAGGAATG 244  
 QY 122 TTGCATTTGATTCCTGAGACATGGAAGAGAGTGTGAGACCAAAACTTG 181  
 DB 245 TAAGCATTTTAAATTCATAGAGAGTGAATCAAAATCTGATCTTAATGCCAAAATG 304  
 QY 182 AGACCTACAAAATGCTGATGTTGCTGCTGAGTCTAATCCTCAGGNAATGATGAAC 241  
 DB 305 AATCATATGACGAGAGATGATGCTAATCTTACCTAATAAATGAGATGATC 364  
 QY 242 CCTACACTGNGCAGATGGCAACTGTGGCGA 272  
 DB 365 CCTATACACTCAATATGAGAAATGTGGAGA 395

RESULT 11  
 AF047838 3022 bp mRNA ROD 26-NOV-1998  
 LOCUS Mus musculus calcium-sensitive chloride conductance protein-1  
 DEFINITION (mClca1) mRNA, complete cds.  
 ACCESSION AF047838  
 NID 93925280  
 VERSION AF047838.1 GI:3925280  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 3022)  
 AUTHORS Gandhi,R., Elble,R.C., Gruber,A.D., Schreuer,K.D., Ji,H.L.,  
 Fuller,C.M. and Pauli,B.U.  
 TITLE Molecular and functional characterization of a calcium-sensitive

chloride channel from mouse lung  
JOURNAL U. Biol. Chem. 273 (48), 32096-32101 (1998)  
MEDLINE 99041960  
REFERENCE 2 (bases 1 to 3022)  
AUTHORS Gandhi, R., Eblen, R.C., Gruber, A.D. and Paul, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (12-FEB-1998) Vet. Pathology, Cornell Univ., WVC C4-146,  
Ithaca, NY 14856, USA

FEATURES  
Source Location/Qualifiers  
1..3022  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/tissue\_type="whole lung"  
1..3022  
/gene="mclca1"  
/gene="mclca1"  
/note="calcium-activated chloride channel: Alternative  
chloride channel to CFTR that may influence severity of  
Cystic Fibrosis in different tissues; mCAC-1"  
/product="calcium-sensitive chloride conductance  
protein-1"  
/protein\_id="AAC79982.1"  
/db\_xref="PID:g3925281"  
/db\_xref="GI:3925281"  
/translation="MPELOVLEFLTLHLNONTSSKVNLSNGEYVIAINPSVE  
DERLIPSTIKENVOASTYLEFASOGRYENISILVPMTKSKPEYLMKRESYDKAD  
VIVADPHLOHGDDEYTLQYGGCGRGVYIHTPEFLTLNLRITGRGRVFEVMAHL  
RMGVDEYNVQPFYMSKRTIEATRCSTRTINVANCERNGCVTRACRDSKTRL  
YEKCTFIPDKIOTAGASIMENONLSVECTEKNHNAEAPLQNMCKNRSTWYI  
KTSADFNAPMGTETAPPEPTFLSKSRVCLVLDKSGMDKEDRLRMNAAL  
YLQIVKRESWGLVTPDSNAHIONYLKITSDDYOKITANLEQASGGTSTCHGLQ  
AGPATSSDOSTSGSEIVLITDGEDNIRCFVRSQALITIALGPSAAELELT  
LDSVTGDRFVITWYMKPEIILQDPKGRKITSDFODDLNRSARLQIPGTAETG  
LTVYSTIGTSQILITVTRAPSPMEPLATAMHOSQTAQPSRMIVARVQSGFL  
PVLGANTALIEAHEHGOVLELMDNGADTVKNDGITYRTFTDYGNGVSLKAVY  
QAOBKRTLSLRQKNSLYIPGVENKIVNPRPVQEALEATVEDERVTSKGS  
FTVSGAPPDGDHARVFPSPKVTDLAEFTIGYIHLTWTPKGVLDNGRAHRYIIRMSO  
HPDLQEDFNATLVNASSLIPKESKTEFKPEFKPEFKIANGIOLYIAIQADNEAST  
TSFVSNIAQVAKLTSLSDSISALGDLSAISMRTIMGLTVIFNSILN"

BASE COUNT 923 a 710 c 661 g 728 t  
ORIGIN  
Query Match 41.5% Score 113.4; DB 12; Length 3022;  
Best Local Similarity 63.1% Pred. No. 2.2e-20;  
Matches 171; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
2 TTGCAATCGACCCCAATGCGAGAGATGAACACCTCATCAACAATAAGACATGG 61  
Db 124 TTGCCATTAAACCCAGTGTCCGGAGGAGCAAGGCTATCCAGCATTAAGGAATGG 183  
QY 62 TGACCCAGGATCTCTGTATCTGTTGAAGCTACAGAAAGGATTTATTCAAAATG 121  
Db 184 TAACGCAAGCTCTACACCTGTTGAAGCCAGCAAGAAAGATTTATTCAAAACA 243  
QY 122 TTGCCATTATTCTCTAATCAATGGAAGCAAGAGTACTATGTGACCAAAATCTG 181  
Db 244 TAACGATATTGTGCTCCGATGACCTGGAAGTCAAAACCTGACTATTAAGCAAAAGAG 303  
QY 182 AGACCTACAAAATGCTGATGTTGCTGAGTCTANTCCCTCCAGGNAATATGATGAC 241  
Db 304 AATGATATGACAAAGACAGACGCTCATAGTTGGGATCTCCTGCAACATGGAAGACGC 363  
QY 242 CCTACACTGNCAGATGGCAACTGTGGCGA 272  
Db 364 CCTAACCCCTCAGATATGGACAGATGGGGA 394

RESULT 12  
AF052746

LOCUS AF052746 3471 bp mRNA ROD 08-APR-1999  
DEFINITION Mus musculus chloride channel Cacc mRNA, complete cds.  
ACCESSION AF052746  
VERSION 93560546  
KEYWORDS AF052746.1 GI:3560546  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3471)  
AUTHORS Romio, L., Musante, L., Cinti, R., Seri, M., Moran, O., Zegarra-Moran, O.  
and Galletta, L.J.V.  
TITLE Characterization of a murine gene homologous to the bovine Cacc  
chloride channel  
JOURNAL Gene 228 (1-2), 181-188 (1999)  
MEDLINE 99173882  
REFERENCE 2 (bases 1 to 3471)  
AUTHORS Romio, L., Musante, L., Cinti, R., Moran, O., Seri, M. and  
Galletta, L.J.V.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto  
Giannina Gaslini, Genova I-16148, Italy

FEATURES  
Source Location/Qualifiers  
1..3471  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/db\_xref="dbEST:AA245677"  
47..2752  
/product="chloride channel Cacc"  
/protein\_id="AAC35003.1"  
/db\_xref="PID:g3560547"  
/db\_xref="GI:3560547"  
/translation="MPELOVLEFLTLHLNONTSSKVNLSNGEYVIAINPSVE  
DERLIPSTIKENVOASTYLEFASOGRYENISILVPMTKSKPEYLMKRESYDKAD  
VIVADPHLOHGDDEYTLQYGGCGRGVYIHTPEFLTLNLRITGRGRVFEVMAHL  
RMGVDEYNVQPFYMSKRTIEATRCSTRTINVANCERNGCVTRACRDSKTRL  
YEKCTFIPDKIOTAGASIMENONLSVECTEKNHNAEAPLQNMCKNRSTWYI  
KTSADFNAPMGTETAPPEPTFLSKSRVCLVLDKSGMDKEDRLRMNAAL  
YLQIVKRESWGLVTPDSNAHIONYLKITSDDYOKITANLEQASGGTSTCHGLQ  
AGPATSSDOSTSGSEIVLITDGEDNIRCFVRSQALITIALGPSAAELELT  
LDSVTGDRFVITWYMKPEIILQDPKGRKITSDFODDLNRSARLQIPGTAETG  
LTVYSTIGTSQILITVTRAPSPMEPLATAMHOSQTAQPSRMIVARVQSGFL  
PVLGANTALIEAHEHGOVLELMDNGADTVKNDGITYRTFTDYGNGVSLKAVY  
QAOBKRTLSLRQKNSLYIPGVENKIVNPRPVQEALEATVEDERVTSKGS  
FTVSGAPPDGDHARVFPSPKVTDLAEFTIGYIHLTWTPKGVLDNGRAHRYIIRMSO  
HPDLQEDFNATLVNASSLIPKESKTEFKPEFKPEFKIANGIOLYIAIQADNEAST  
TSFVSNIAQVAKLTSLSDSISALGDLSAISMRTIMGLTVIFNSILN"

BASE COUNT 1096 a 777 c 736 g 862 t  
ORIGIN  
Query Match 41.5% Score 113.4; DB 12; Length 3471;  
Best Local Similarity 63.1% Pred. No. 2.2e-20;  
Matches 171; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
2 TTGCAATCGACCCCAATGCGAGAGATGAACACCTCATCAACAATAAGACATGG 61  
Db 153 TTGCCATTAAACCCAGTGTCCGGAGGAGCAAGGCTCATCCAGCATTAAGGAATGG 212  
QY 62 TGACCCAGGATCTCTGTATCTGTTGAAGCTACAGAAAGGATTTATTCAAAATG 121  
Db 213 TAACGCAAGCTCTACACCTGTTGAAGCCAGCAAGAAAGATTTATTCAAAACA 272  
QY 122 TTGCCATTATTCTCTAATCAATGGAAGCAAGAGTACTATGTGACCAAAATCTG 181  
Db 273 TAACGATATTGTGCTCCGATGACCTGGAAGTCAAAACCTGACTATTAAGCAAAAGAG 332  
QY 242 CCTACACTGNCAGATGGCAACTGTGGCGA 272  
Db 333 AATGATATGACAAAGACAGACGCTCATAGTTGGGATCTCCTGCAACATGGAAGACGC 392

RESULT 12  
AF052746

QY 242 CCTACACTGCGAGATGGCACTGTGGCA 272  
||||| 111 111 111 111 111 111  
Db 393 CCTACACCTTCAGTATGACAGTGGGA 423

RESULT 13  
AB026833 3604 bp mRNA PRI 26-MAY-1999  
LOCUS Homo sapiens mRNA for chloride channel protein, complete cds.  
AB026833  
ACCESSION 94887600  
NID AB026833.1 GI:4887600  
VERSION chloride channel protein.  
KEYWORDS Homo sapiens adult corneal epithelium cDNA to mRNA,  
SOURCE clone\_1lb:lambda ZAP clone:lambda 13802.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Itoh, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.  
TITLE Isolation and characterization of chloride channel in human corneal  
JOURNAL epithelium  
REFERENCE Unpublished (1999)  
AUTHORS 2 (bases 1 to 3604)  
DIRECT SUBMISSION Itoh, R., Kawamoto, S. and Okubo, K.  
TITLE Submitted (28-APR-1999) to the DDBJ/EMBL/Genbank databases. Reiko  
JOURNAL Itoh, Institute for Molecular and Cellular Biology, Osaka  
University, Yamada-oka 1-3, Suita, Osaka 565-0871, Japan  
(E-mail:reiko@imcb.osaka-u.ac.jp, Tel:81-6-6879-7992,  
Fax:81-6-6877-1922)

FEATURES  
source  
1. 3604  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone\_1lb="lambda 13802"  
/clone\_1lb="lambda ZAP"  
/dev\_stage="adult"  
/map\_1p32"  
/tissue\_type="corneal epithelium"  
42..2873  
/codon\_start=1  
/product="chloride channel protein"  
/protein\_id="BA77810.1"  
/db\_xref="pid:d1041578"  
/db\_xref="pid:94887601"  
/db\_xref="GI:4887601"

CDS  
1  
/translation="MTORSIAGPICNLKVTLLVALSSLPFLGAGVQLQDNKGYLL  
IAINPOVENQNLISIKEMITEASYLLNATKRREFFENIKILIPATRKANNRSIK  
QSYEKANIVTDWYGAGHDPTTLOYRCGEGKIHPTPLNDNLTAAGSKGR  
VVEHAWLRMCVDEYINDKPEYINGOQIKVTCSSDITGIFVEKPCQENCI  
SVLEKGCETFINSTONATASIMFOSLSVVEFCASHTNEAPLEKQMSLRAM  
DYTTSADPHSEFPMKNCETLPPPTSLVQADGKVCYLVDSSKMAEDRLIQOA  
AFTLMQIYEHTFPGIASFDSKGERALDHOINSDDKILYSITPTYSKTDISI  
CSGLKGEFVEKLNKAGSVILVTSDDKILGCLPTVLSSGSTHISIALGSKR  
PMLELSRLTGLKFEFSDPISNSNSMIDFSGTGIPOHIOLESTGENVSPH  
QIKNTVTNDNTGNDTMTLWQASGPRPIIFDPGRKYVTNNFTNTLFTASIMI  
PCTAKGHWYTLNTHSLQALKVTTSRASNVAVPATVAFVRSDLHPHPMI  
YANVKGFEPIINATVTATPEPTGDPVTLRLDGAADVAKNDGITSRYEFSFAM  
GRSLKRVHNHSPSISTPAHSIPGSHAMVYPTGTANGNOMAPKRSVNRDEERWG  
FSRVSGGSFVGVGAPGHPDVPFPCKLIDLEAVVEELTSTWAPEDDQOAT  
SEIRMSKSLQNIQDQFNALIVNTSKRPQAGIREITFPQISTNPEHQNET  
HESHRIYVAIRAMDNSLOSASVNIQAFLFIPNSDPVPARDYLLIKGVLTAAGLIG  
ICLIIVVHTHSRRKRRADKRENGPKLL"

BASE COUNT 1113 a 740 c 732 g 1017 t 2 others

Query Match 27.0%; Score 73.6; DB 9; Length 3604;  
Best Local Similarity 56.7%; Pred. No. 6.2e-10;  
Matches 152; Conservative 0; Mismatches 113; Indels 3; Gaps 1;

QY 2 TTGCAATCGACCCCATGTGCCAGATGAACACTCATTCACAAATTAAGCATG 61  
||||| 111 111 111 111 111 111 111 111 111 111 111 111  
Db 175 TTGCCATTAATCTCTCAGGTACTGAGAAATCAGAACTTCATCAACATTAAAGAAATGA 234

QY 62 TGACCCAGCATCTCTGTATCTGTGTTGAAGCTACGAAAGCATTTTATTCAAAATG 121  
||||| 111 111 111 111 111 111 111 111 111 111 111 111  
Db 235 TAACGTAGCTTATTTTACTTATTTATGCTACCAAGAGAGATATTTTTCAGAAATA 294

QY 122 TTGCATTTTGTATCTCTGAAACATGGAAGCAAAAGNTGACTATGTGAGACCAAACTTG 181  
||||| 111 111 111 111 111 111 111 111 111 111 111 111  
Db 295 TAAAGATTTTAAATCTCTGACCATGGAAGCAAGCTATATATAC---ACCAAAATTAACAAG 351

QY 182 AGACCTACAAATAATGCTGATGCTGTGCTGCTAGCTANTCTCCAGNAATAGAAC 241  
||||| 111 111 111 111 111 111 111 111 111 111 111 111  
Db 352 AATCAATATGAAAGGCAAAAGTATGACTGCTGATGAGGCAATGAGATGATC 411

QY 242 CCTACACTGCGAGATGGCACTGTGG 269  
||||| 111 111 111 111 111 111 111 111 111 111 111 111  
Db 412 CATACACCTTCACAAATACAGAGGTGTGG 439

RESULT 14  
AC004749 60791 bp DNA PRI 27-MAY-1998  
LOCUS Homo sapiens chromosome 5, P1 clone 632A8 (LBNL H29), complete  
DEFINITION sequence.  
ACCESSION AC004749 L49048 L77057 L77056 L77055 L48484 L77054 L77053 L81380  
L81379 L81798 AC001489 AC000975 L81797 L81796 AC000976 AC000974  
NID AC002208 AC000977 L81795 L81628 L78759 L78758  
93157918  
VERSION AC004749.1 GI:3157918  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 60791)  
AUTHORS Kimerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,  
Davis, C.A., Kadner, K., Miguel, T., Pfluck, S., Pollard, M.,  
Rojeski, H., Subramanian, S. and Martin, C.H.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAY-1998) Human Genome Center, DOE Joint Genome  
Institute, Lawrence Berkeley National Laboratory, MS 74-157,  
Berkeley, CA 94720, U.S.A.  
COMMENT Sequence submitted by:  
DOE Joint Genome Institute.  
LOCATION/Qualifiers  
1. 60791  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="5q"  
/clone="632A8"  
/chromosome="5"  
/note="LBNL H29"  
1..60791  
/note="(T)27"  
/rpt\_type=tandem  
/rpt\_unit=T  
453..757  
/rpt\_family="MERIB"  
complement(1157..1283)

```
misc_feature /note="GRAIL 2 excellent exon, frame 0"  
complement(1177..1282)  
/note="83% identity mouse mp17d11.r1"  
/db_xref="dbEST:AA288684"  
repeat_region complement(1882..2044)  
/rpt_family="Alu"  
1891..1913  
/note="(T)23"  
/rpt_type=tandem  
/rpt_unit=T  
repeat_region 5204..5503  
/rpt_family="Alu"  
6376..6912  
/rpt_family="Alu"  
8028..8115  
/note="GRAIL 2 excellent exon, frame 2"  
complement(9142..9288)  
/rpt_family="MER5"  
9356..9674  
/rpt_family="Alu"  
9611..10326  
/rpt_family="L1"  
repeat_region complement(10569..10869)  
/rpt_family="Alu"  
10859..10943  
/note="GRAIL 2 excellent exon, frame 1"  
complement(11133..11417)  
/note="99% identity EST60905"  
/db_xref="dbEST:AA352848"  
repeat_region complement(12263..12538)  
/rpt_family="Alu"  
12276..12296  
/note="(T)21"  
/rpt_type=tandem  
/rpt_unit=T  
repeat_region complement(14099..14403)  
/rpt_family="Alu"  
14892..15199  
/note="97% identity EST02082"  
/db_xref="dbEST:M85566"  
16203..16362  
/rpt_family="Alu"  
17932..18067  
/rpt_family="MER42"  
repeat_region complement(18549..18857)  
/rpt_family="Alu"  
18552..18868  
/note="(T)21"  
/rpt_type=tandem  
/rpt_unit=T  
repeat_region complement(19792..20110)  
/rpt_family="Alu"  
repeat_region complement(20826..21152)  
/rpt_family="Alu"  
21162..21549  
/note="99% identity YU10e03.r1"  
/db_xref="dbEST:H79907"  
repeat_region complement(21500..21824)  
/rpt_family="Alu"  
22829..23118  
/rpt_family="Alu"  
23149..23457  
/rpt_family="Alu"  
repeat_region complement(23659..23973)  
/rpt_family="L1"  
24620..25456  
/rpt_family="Alu"  
25651..26710  
/rpt_family="Alu"  
25933..26038  
/rpt_family="Alu"  
26390..26695  
/rpt_family="Alu"  
repeat_region  
repeat_region complement(26796..26904)  
/rpt_family="MIR"  
complement(27505..27616)  
/note="99% identity"  
/db_xref="dbEST:D63193"  
repeat_region complement(29807..30188)  
/rpt_family="Alu"  
complement(31287..31532)  
/rpt_family="Alu"  
31649..31828  
/rpt_family="MER46"  
31831..31976  
/rpt_family="Alu"  
repeat_region complement(32558..32880)  
/rpt_family="Alu"  
complement(32887..33189)  
/rpt_family="Alu"  
33219..33558  
/rpt_family="Alu"  
33559..33843  
/rpt_family="Alu"  
repeat_region complement(34684..34783)  
/rpt_family="L1"  
35259..35558  
/rpt_family="Alu"  
35361..35867  
/rpt_family="Alu"  
36176..36203  
/note="(AC)14"  
/rpt_type=tandem  
/rpt_unit=AC  
repeat_region complement(37312..42768)  
/rpt_family="L1"  
complement(43077..43389)  
/rpt_family="Alu"  
43834..44199  
/rpt_family="Alu"  
44128..44446  
/rpt_family="Alu"  
44424..44448  
/note="(A)20"  
/rpt_type=tandem  
/rpt_unit=A  
repeat_region complement(44448..44837)  
/rpt_family="Alu"  
complement(44895..45025)  
/rpt_family="Alu"  
45828..46140  
/rpt_family="Alu"  
complement(46222..46303)  
/rpt_family="MER42C"  
46466..46636  
/rpt_family="MER44C"  
complement(46987..47304)  
/rpt_family="Alu"  
complement(47536..47862)  
/rpt_family="Alu"  
48111..48231  
/rpt_family="MIR"  
48705..54852  
/rpt_family="L1"  
complement(55742..56085)  
/rpt_family="Alu"  
55940..55959  
/note="(T)20"  
/rpt_type=tandem  
/rpt_unit=T  
repeat_region complement(56109..56261)  
/rpt_family="MER20"  
complement(56361..56800)  
/note="97% identity YP96g11.r1"  
/db_xref="dbEST:R92125"  
56771..56875  
repeat_region
```

```
repeat_region      /rpt_family="MLTIF"
                    complement(56884..57190)
repeat_region      /rpt_family="Alu"
                    58783..58908
                    /rpt_family="MSTC"

Query Match
Best Local Similarity 13.7%: Score 37.4; DB 11; Length 60791;
Matches 95; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 12 CCCCAATGTCAGAGATGAAAGACTTCAACAATTAAGACATGTCAGCCAGGC 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 58041 CCCGCAAGGACACACACAATATGACACATCTATACATTAAGTCTGCAAGAGAT 58100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 72 ATCTGTGATCTGTTTGAAGCTACAGAGAAAGCATTTTATTTCAAAAATGTCGCAATTT 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 58101 GACTCACTAACACAACTATGCTCATCTGACATATTTAGTGAATGATCATCCAGAAAT 58160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 132 GATTCCTGAACATGAGACAAAGAGTGTGAGACGACAAAGTGTGAGACCTACAA 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 58161 GTGTCCCTGAAGAGTCAATGGAATATAGTATATGAAAGCCAAACAGAAACCATGAA 58220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 192 AATGCTGATGT 203
    ||| ||| ||| |||
DB 58221 AAGGATCAACT 58232

RESULT 15
GDCOL6A1A 21180 bp DNA VRT 19-NOV-1992
LOCUS G.domesticus gene for collagen alpha 1 type VI.
ACCESSION X57898
VERSION X57998.1 GI:62874
NID 962874
KEYWORDS cell adhesion activity; collagen alpha 1 type VI; extracellular
SOURCE matrix protein; transformation-sensitive protein.
ORGANISM chicken.
Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 21180)
JOURNAL Submitted (22-FEB-1991) B. Trueb, Swiss Federal Inst of Technology,
BIOCHEMISTRY I, ETH Zentrum, CH-8092 Zurich, Switzerland
2 (bases 1 to 21180)
AUTHORS Walchli,C., Koller,E., Trueb,J. and Trueb,B.
TITLE Structural comparison of the chicken genes for alpha 1(VI) and
alpha 2(VI) collagen
JOURNAL Eur. J. Biochem. 205 (2), 583-589 (1992)
MEDLINE 92241293
FEATURES
source
    location/Qualifiers
    1..21180
    /organism="Gallus gallus"
    /db_xref="taxon:9031"
    /dev_stage="adult"
    /tissue_type="liver"
    /clone_lib="EMBL3"
    1..203
    /number=1
    join(1..203,874..1003,2116..2316,2538..2697,2939..3061,
    3153..3173,3328..3348,3440..3484,3721..3774,4766..4808,
    4906..4932,5075..5101,5204..5248,5414..5467,5884..5946,
    6473..6535,6680..6769,6880..6942,8248..8310,11227..11289,
    11776..11838,12534..12584,12669..12704,13159..13221,
    13576..13641,13787..13822,14471..14507,14935..14943,
    15492..15625,16591..16700,16980..17163,17366..17540,
    18066..18095,18831..20870)
    /gene="Col6A1"
    /number=7
    join(107..203,874..1003,2116..2316,2538..2697,2939..3061,
    3153..3173,3328..3348,3440..3484,3721..3774,4766..4808,
    4906..4932,5075..5101,5204..5248,5414..5467,5884..5946,
    6473..6535,6680..6769,6880..6942,8248..8310,11227..11289,
    11776..11838,12534..12584,12669..12704,13159..13221,
    13576..13641,13787..13822,14471..14507,14935..14943,
    15492..15625,16591..16700,16980..17163,17366..17540,
    18066..18095,18831..20870)
    /gene="Col6A1"
    /number=8
    3485..3720
```



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:22 ; Search time 3059.95 Seconds  
(without alignments)  
175.984 Million cell updates/sec

Title: US-09-049-696-2

Perfect score: 273  
Sequence: 1 GTTGCATGACCCCAATGT.....AGATGGCACTGTGGCGAG 273

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	11.4	428	30	284017	284017 SS284017 Po
2	19	7.0	688	42	A133878	A133878 GH10661.5
3	19	7.0	447	44	A127324	A127324 G14509.x
4	19	7.0	593	46	A155107	A155107 ID21194.3
5	18	6.6	537	22	H10546	H10546 ym04d01.t1
6	18	6.6	518	26	Z71911	Z71911 CHEST809 Go
7	18	6.6	490	27	AA002184	AA002184 zh81908.s
8	18	6.6	517	29	AA142874	AA142874 z149b05.s
9	18	6.6	322	29	AA142895	AA142895 z149f05.s
10	18	6.6	274	31	AA298669	AA298669 EST114502
11	18	6.6	341	35	AA542676	AA542676 fa06e01.r
12	18	6.6	398	43	AA669224	AA669224 ac16b04.s
13	18	6.6	707	45	AI241467	AI241467 qh69a11.x
14	18	6.6	429	48	AU005990	AU005990 AU005990
15	18	6.6	193	22	R43122	A1558650 fb68h03.y
16	17	6.2	483	22	R55842	R43122 y910d04.s1
17	17	6.2	489	22	R75771	R55842 y989d02.t1
18	17	6.2	291	23	F14976	R75771 y121a10.r1
19	17	6.2	344	24	D18308	F14976 SSC7H01 Por
20	17	6.2	537	24	N28457	D18308 MUGS01184
21	17	6.2	341	25	N72239	N28457 yx35g05.r1
22	17	6.2	533	26	W97675	N72239 z401f07.s1
23	17	6.2	421	27	AA001230	W97675 m198b01.r1
24	17	6.2	535	27	W99845	AA001230 zh82g03.r
25	17	6.2	629	29	AA142696	W99845 m928h11.r1
26	17	6.2	570	29	AA187968	AA142696 mr86e10.r
27	17	6.2	744	30	AA198886	AA187968 zp62a03.r
28	17	6.2	429	30	AA220968	AA198886 m16g02.i
29	17	6.2	475	31	AA284073	AA220968 zt01g01.r
30	17	6.2	319	32	AA334806	AA284073 zs49a11.s
31	17	6.2	409	33	AA412477	AA334806 EST139072
32	17	6.2	450	34	AA510978	AA412477 zt95g10.s
33	17	6.2	466	36	AA606193	AA510978 vg38a09.r
34	17	6.2	386	37	AA697023	AA606193 fa16f06.s
35	17	6.2	454	39	AA887216	AA697023 GM08947.5
36	17	6.2	413	40	AA978118	AA887216 o11b02.s
37	17	6.2	376	41	AT019560	AA978118 o133c03.s
38	17	6.2	376	41	AT019560	AA978118 op33d04.s
39	17	6.2	414	41	AT019560	AA978118 ua90d02.r
40	17	6.2	376	41	AT019560	AT019560 ua91g08.r
41	17	6.2	724	45	A1324543	AT019560 mr86e10.y
42	17	6.2	423	45	A1356791	A1324543 qy17g03.x
43	17	6.2	536	46	A1416317	A1356791 fb20h06.y
44	17	6.2	578	50	A1657526	A1416317 LD35171.3
45	17	6.2	578	50	A1657526	A1455641 fc41d08.x

#### ALIGNMENTS

RESULT	1					
LOCUS	284017	428 bp	MRNA	EST	30-JAN-1997	
DEFINITION	SS284017	Porcine small intestine cDNA library	Sus scrofa	CDNA clone		
ACCESSION	284017	G12803 5', mRNA sequence.				
NID	91806336					
VERSION	284017.1	GI:1806336				

```

KEYWORDS      EST.
SOURCE         pig.
ORGANISM       Sus scrofa
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE      1 (bases 1 to 428)
AUTHORS        Wintero,A.K., Fredholm,M. and Davies,W.
TITLE          Evaluation and characterization of a porcine small intestine cDNA
               library: analysis of 839 clones
JOURNAL        Mamm. Genome 7 (7), 509-517 (1996)
MEDLINE        96327607
COMMENT        On Sep 12, 1996 this sequence version replaced gi:1404772.

FEATURES
  source       Contact: A.K. Wintero
               Department of Animal Science and Animal Health, Division of Animal
               Genetics, The Royal Veterinary and Agricultural University
               Bulowsvej 13, 1870 Frederiksberg C, Denmark.
               Location/Qualifiers
               1..428
               /organism="Sus scrofa"
               /db_xref="taxon:9823"
               /clone="c12a03"
               /clone_lib="Porcine small intestine cDNA library"
               /note="directionally cloned cDNA in X1-blue MRF"

BASE COUNT    134 a      87 c      100 g      102 t      5 others

ORIGIN
Query Match      11.4%; Score 31; DB 30; Length 428;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 TATTTCAAAATGTGGCATTTCATCCG 139
      |||||||||||||||||||||||||||
DB 242 TATTTCAAAATGTGGCATTTCATCCG 272

RESULT 2
LOCUS      A1133878      688 bp      mRNA      EST      18-NOV-1998
DEFINITION GH10661.5prime GH Drosophila melanogaster head pot2 Drosophila
ACCESSION  A1133878
NID        93626436
VERSION    A1133878.1 GI:3626436
KEYWORDS   EST.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
           Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
           Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
           Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
           1 (bases 1 to 688)
           Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
           Brokstein,P., Lewis,S. and Rubin,G.M.
           BDGP/HMT Drosophila EST Project
           Unpublished (1997)

REFERENCE   1
AUTHORS     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST. est@fruitfly.berkeley.edu
            Plate: 106 row: F column: 1
            High quality sequence stop: 529.
            Location/Qualifiers
            1..688
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="GH10661"
            /clone_lib="GH Drosophila melanogaster head pot2"
            /sex="male and female"
            /dev_stage="adult"

FEATURES
  source

```

```

/1ab_host="DH5 - alpha"
/note="Organ: head; Vector: pot2; site_1: EcoRI; site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."

BASE COUNT    201 a      178 c      131 g      178 t

ORIGIN
Query Match      7.0%; Score 19; DB 42; Length 688;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 ATGAACACATCATTACACA 47
      |||||||||||||||||||
DB 240 ATGAACACATCATTACACA 258

RESULT 3
LOCUS      A1273224/c      447 bp      mRNA      EST      29-JAN-1999
DEFINITION g145a09.x1 NCI-CGAP_C08 Homo sapiens cDNA clone IMAGE:1875256 3'
ACCESSION  A1273224
NID        93895492
VERSION    A1273224.1 GI:3895492
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Homiidae; Homo.
           1 (bases 1 to 447)
           NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
           National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
           Unpublished (1997)
           On May 1, 1998 this sequence version replaced gi:2761092.

JOURNAL
COMMENT

```

```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
Clone Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNW at:
www.bio.lnl.gov/bbrp/image/image.html

Insert Length: 777 Std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 360.
Location/Qualifiers
1..447
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1875256"
/clone_lib="NCI-CGAP_C08"
/tissue_type="adeno carcinoma"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT    127 a      90 c      78 g      152 t

ORIGIN

```



Query Match 7.0%; Score 19; DB 44; Length 447;  
 Best Local Similarity 100.0%; Pred. No. 4;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 107 TTTATTTCAAAATGTTGC 125  
 ||||||||||||||||  
 DB 416 TTTATTTCAAAATGTTGC 398

RESULT 4  
 A1455107/c 593 bp mRNA EST 09-MAR-1999  
 LOCUS LD21194.3prime LD Drosophila melanogaster embryo POT2 Drosophila  
 melanogaster cDNA clone LD21194 3prime, mRNA sequence.  
 ACCESSION A1455107  
 NID 94299645  
 VERSION A1455107.1 GI:4299645  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 593)  
 Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,  
 Brokstein, P., Lewis, S. and Rubin, G.M.  
 BDGP/HMI Drosophila EST Project  
 Unpublished (1997)  
 On May 7, 1998 this sequence version replaced gi:3121029.  
 Other ESTs: LD21194.5prime  
 Contact: Harvey, D.  
 G. M. Rubin-Molecular and Cell Biology  
 University of California Berkeley  
 539 USA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: <http://www.fruitfly.org/EST/estfruitfly.berkeley.edu>  
 Based upon the presence of a XhoI site followed by a run of 14 or  
 more T residues at the beginning of the sequence, this clone was  
 polyadenylated. The resulting Poly-T sequence has been removed.  
 Plate: 211 row: H column: 10  
 High quality sequence stop: 498.  
 Location/Qualifiers  
 1..593  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="LD21194"  
 /clone\_1lb="LD Drosophila melanogaster embryo POT2"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="XLI Blue"  
 /note="Organ: embryo; Vector: POT2; Site\_1: EcoRI; Site\_2:  
 XhoI; Sized fractionated cDNAs were directly ligated into  
 POT2."

BASE COUNT 152 a 104 c 145 g 192 t  
 ORIGIN

Query Match 7.0%; Score 19; DB 46; Length 593;  
 Best Local Similarity 100.0%; Pred. No. 3.6;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 ATGAACACTCATTCACA 47  
 ||||||||||||||||  
 DB 453 ATGAACACTCATTCACA 435

RESULT 5  
 H10546 557 bp mRNA EST 23-JUN-1995  
 LOCUS H10546  
 DEFINITION ym04d01.r1 Soares infant brain INIB Homo sapiens cDNA clone  
 IMAGE:46959 5; similar to SP:KPM.HUMAN P11309 PIM-1 PROTO-ONCOGENE  
 SERINE/THREONINE-PROTEIN KINASE;; mRNA sequence.  
 ACCESSION H10546

NID 9875368  
 VERSION H10546.1 GI:875368  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 557)  
 Hillier, L., Clark, N., Dubague, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)  
 On May 10, 1995 this sequence version replaced gi:805969.

TITLE  
 JOURNAL  
 COMMENT

Contact: Wilson, R.K.  
 Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
 Insert Size: 1866  
 High quality sequence stops: 400  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Insert Length: 1866 Std Error: 0.00  
 Seq primer: M13Rpl  
 High quality sequence stop: 400.  
 Location/Qualifiers  
 1..557  
 /organism="Homo sapiens"  
 /db\_xref="GDB:419500"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:46959"  
 /clone\_1lb="Soares infant brain INIB"  
 /sex="female"  
 /dev\_stage="73 days post natal"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: whole brain; Vector: Lambda BA; Site\_1: Not  
 I; Site\_2: Hind III; 1st strand cDNA was primed with a Not  
 I - 01590(dT) primer [5'  
 AACTGAGAGAAATCGCGCCGACGAGAAATTTTTTTTTTTT 3'];  
 double-stranded cDNA was ligated to Hind III adaptors  
 (Pharmacia), digested with Not I and directionally cloned  
 into the Not I and Hind III sites of the Lambda BA vector.  
 Library went through one round of normalization. Library  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 150 c 152 g 141 t  
 ORIGIN

Query Match 6.6%; Score 18; DB 22; Length 557;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 233 ATGATGACCTTACTG 250  
 ||||||||||||||||  
 DB 351 ATGATGACCTTACTG 368

RESULT 6  
 271911 518 bp mRNA EST 13-NOV-1996  
 LOCUS CHEST809 Goat mammary gland Capra hircus cDNA clone EST8-09; mRNA  
 sequence.  
 ACCESSION 271911  
 NID 91292836  
 VERSION 271911.1 GI:1292836  
 KEYWORDS EST.  
 SOURCE goat.

ORGANISM      Capra hircus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae;  
REFERENCE      1 (bases 1 to 518)  
AUTHORS      Le Provost, F., Lepingle, A. and Martin, P.  
TITLE      A survey of the goat genome transcribed in the lactating mammary  
                 gland  
JOURNAL      Mamm. Genome 7 (9), 657-666 (1996)  
MEDLINE      96359152  
COMMENT      On Apr 14, 1993 this sequence version replaced gi:785308.

FEATURES  
    Source  
            Contact: Le Provost, F.  
            Laboratoire de Genetique Biochimique  
            INRA-CRI  
            78352 Jouy-en-Josas Cedex, FRANCE.  
            Location/Qualifiers  
            1..518  
            /organism="Capra hircus"  
            /db\_xref="taxon:9925"  
            /clone="EST8-09"  
            /clone\_lib="Goat mammary gland"  
            /note="Mammary gland of Capra hircus (goat) "

BASE COUNT      161 a      95 c      94 g      159 t      9 others  
ORIGIN

Query Match      6.6%; Score 18; DB 26; Length 518;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

0x    193 AATGCTGATGTTCTGTT 210  
      |||||  
Db     61 AATGCTGATGTTCTGTT 44

RESULT 7  
AA002184  
LOCUS      AA002184  
DEFINITION      AA002184 490 bp mRNA EST 07-MAY-1997  
                 zhs1g08.s1 Soares\_fetal\_liver.spleen\_1NFUS\_S1 Homo sapiens CDNA  
                 clone IMAGE:427742.3' similar to contains Alu repetitive  
                 element; contains element LTR10 repetitive element ;, mRNA sequence  
ACCESSION      AA002184  
NID      91445099  
VERSION      AA002184.1 GI:1445099  
KEYWORDS      EST.  
SOURCE      human.  
ORGANISM      Homo sapiens  
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
            Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE      1 (bases 1 to 490)  
AUTHORS      Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B.,  
            Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
            Hawkins, M., Hultman, M., Kneba, T., Lacy, M., Le, M., Le, N.,  
            Mawds, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
            Roifling, T., Schellenberg, K., Soares, M.B., Tan, F., Thery, W., J.,  
            Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
            and Marra, M.  
TITLE      Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL      Genome Res. 6 (9), 807-828 (1996)  
MEDLINE      97044478  
COMMENT      On Sep 21, 1992 this sequence version replaced gi:279097.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 680 Std Error: 0.00  
Seq primer: mob.NEGA+ET

FEATURES	High quality sequence stop: 425.
SOURCE	Location/Qualifiers
	1. 490
	/organism="Homo sapiens"
	/db_xref="GDB:1327511"
	/db_xref="taxon:9606"
	/clone="IMAGE:427742"
	/clone_1lb="Soares_fetal_liver_spleen_1NFLS-S1"
	/sex="male"
	/dev_stage="20 week-post conception fetus"
	/lab_host="DH10B (ampicillin resistant)"
	/note="Organ: Liver and Spleen; Vector: p773D (Pharmacla) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen 1NFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer 15', AACGCGAAGATTAATTAAGACTCTTTTCTTTTCTTTT 3'). double-stranded cDNA was ligated to Eco RI adaptors (Pharmacla), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT	176 a 86 c 91 g 135 t 2 others
ORIGIN	
Query Match	6.6%; Score 18; DB 27; Length 490;
Best Local Similarity	100.0%; Pred. No. 13;
Matches	18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	186 CTACAAAATGCTGATGT 203
Db	183 CTACAAAATGCTGATGT 200
RESULT	8
LOCUS	AA142874
DEFINITION	219BD05.81 Soares_pregnant_uterus_NBHPV Homo sapiens cDNA clone IMAGE:505233 3', mRNA sequence.
ACCESSION	AA142874
NID	91712260
VERSION	AA142874.1 GI:1712260
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 517)
AUTHORS	Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chippelli,B., Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Wadls,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaakis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE	Generation and analysis of 280,000 human expressed sequence tags
JOURNAL	Genome Res. 6 (9), 807-828 (1996)
MEDLINE	97044478
COMMENT	On Nov 29, 1993 this sequence version replaced gi:503227.
	Contact: Wilson RK
	Washington University School of Medicine
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
	Tel: 314 286 1800
	Fax: 314 286 1810
	Email: est@watson.wustl.edu
	This clone is available royalty-free through LNL; contact the
	IMAGE Consortium (info@image.lnl.gov) for further information.
	Insert length: 682 Std Error: 0.00
	Seq primer: -40M13 fwd. from Amersham
	High quality sequence stop: 356.
	Location/Qualifiers
FEATURES	



Index (<http://www.tigr.org/tdb/ngi/ngi.html>)  
Seq primer: M13 Reverse.

# FEATURES

Location/Qualifiers

1..274  
/organism="Homo sapiens"  
/db\_xref="ATCC (Inhost):190362"  
/db\_xref="taxon:9606"  
/map="19p12-p13.1"  
/clone\_lib="Pancreas tumor I"  
/dev\_stage="adult"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"  
BASE COUNT 58 a 78 c 71 g 65 t 2 others  
ORIGIN

Query Match 6.6%; Score 18; DB 31; Length 274;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 233 ATGATGAACCTACACTG 250  
|||||  
Db 218 ATGATGAACCTACACTG 235

## RESULT 11

AA542676

LOCUS fa06e01.r1 zebrafish ICRFzfls Danio rerio cDNA clone 4J1 5', mRNA  
DEFINITION sequence.

ACCESSION AA542676

NID 92289367

VERSION AA542676.1 GI:2289367

KEYWORDS EST.

SOURCE zebrafish.

ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Cyprinidae; Rasbora; Danio.

## REFERENCE

AUTHORS

1 (bases 1 to 341)  
Clark, M., Lehnach, H., Appel, B., Eisen, J., Johnson, S., Marra, M.,  
Eddy, S., Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G.,  
Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Martin, J., Moore, B.,  
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y.,  
Wyllie, T., Waterston, R. and Wilson, R.  
WashU Zebrafish EST Project  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1404879.

TITLE

JOURNAL

COMMENT

Contact: Steve Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Steve Johnson lab internal ID - P2\_145 NOTE - For this library, the  
cDNA id field represents a position identifier on the original  
cDNA library preparation plate. cDNA Library Preparation: Matthew  
Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by:  
Washington University Genome Sequencing Center Clone Distribution:  
Genetik, Berlin Tel +49 30 84 13 1235  
Seq primer: T7 ET from Amersham  
High quality sequence stop: 121.  
Location/Qualifiers

## FEATURES

source

1..341  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone\_lib="zebrafish ICRFzfls"  
/sex="mixed"  
/tissue\_type="pooled 26-somite embryos"  
/lab\_host="X11-blue MRF"

BASE COUNT 121 a 83 c 83 g 54 t  
ORIGIN

Query Match 6.6%; Score 18; DB 35; Length 341;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 181 GAGACCTACAAAATGCT 198  
|||||  
Db 83 GAGACCTACAAAATGCT 100

## RESULT 12

AA669224

LOCUS ac16b04.s1 Stragene ovary (#937217) Homo sapiens cDNA clone  
DEFINITION IMAGE:856591 3', similar to TR:G765068 G765068  
PROTEIN-SERINE/THREONINE KINASE. [2] TR:G765067 ;, mRNA sequence.

ACCESSION AA669224

NID 92630723

VERSION AA669224.1 GI:2630723

KEYWORDS EST.

SOURCE human.

## REFERENCE

AUTHORS

1 (bases 1 to 415)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,  
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
On Nov 6, 1997 this sequence version replaced gi:932378.

TITLE

JOURNAL

COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 409.  
Location/Qualifiers

## FEATURES

source

1..415  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Stragene ovary (#937217)"  
/sex="female"  
/dev\_stage="49 year Old"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: ovary; Vector: Bluescript SK-; Site\_1: EcoRI;  
Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT.  
Total ovary tissue, normal, caucasian. Average insert



JOURNAL Unpublished (1998)  
On Jun 5, 1998 this sequence version replaced gl:318312.

Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu  
CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
Resourcenet/Uniprimerdatenbank, Berlin, Germany (web address:  
www.rzpd.de)

zebrafish identity (p-value greater than 1e-99) found to:  
gl:2289367/igb/AA542676/AA542676 fa06e01.r1 zebrafish ICRFzfls Danio  
rerio CDNA  
Seq primer: T3 ET from Amersham  
High quality sequence stop: 337.  
Location/Qualifiers

FEATURES

source  
1. 429  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone\_lib="Zebrafish WashU MPING EST"  
/sex="mixed"  
/tissue\_type="26 somite embryos, adult livers, shield  
stage embryos"  
/lab\_host="XLI-blue MRF"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st  
strand CDNA was primed with a Not I - Oligo(dT)15 primer  
15'pGACTGATCTAGATCGCGAGCGCGCCCTTTTCTTTT3';  
double-stranded CDNA was ligated to Sal I adaptors (BRL),  
digested with Not I and cloned into the Not I and Sal I  
sites of the pSPORT1 vector (BRL). Library was constructed  
by Matthew Clark (Lehrach lab, ICRF, London and Max Planck  
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
analysis were selected following oligonucleotide  
hybridization fingerprinting of arrayed clones from  
zebrafish late somitogenesis (26 ss), adult liver or  
embryonic shield stage (5.6 h) libraries. Fingerprint  
data were used to computationally cluster cDNAs, and a  
single CDNA from each cluster was chosen for sequencing.  
In some cases multiple members of the same cluster were  
sequenced to assess clustering parameters or single clones  
were sequenced additional times to assess quality  
control."

BASE COUNT 143 a 113 c 99 g 73 t 1 others  
ORIGIN

Query Match 6.6%; Score 18; DB 48; Length 429;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 181 GAGACCTACAAAATGCT 198  
|||||  
DB 94 GAGACCTACAAAATGCT 111

Search completed: August 6, 1999, 09:22:24  
Job time: 6215 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:32:37 : Search time 3075.15 Seconds  
(without alignments)  
175.114 Million cell updates/sec

Title: US-09-049-696-2

Perfect score: 273  
Sequence: 1 GTTGCATCGACCCCAATGT.....AGATGGCACTGTGGCGAG 273

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	206	75.5	428	30	Z84017	Z84017 SSZ84017 Po
2	113.4	41.5	638	45	A1390317	A1390317 mx03b07.y
3	99.6	36.5	417	30	AA245677	AA245677 mx03b07.r
4	36	13.2	441	20	T09430	T09430 EST07323.in
5	35.4	13.0	265	36	AA645367	AA645367 vs80g05.r
6	34.8	12.7	532	47	A1499625	A1499625 lo03a09.x
7	34.6	12.7	573	42	A1515458	A1515458 qc74e05.x
8	34.2	12.5	287	20	T38551	T38551 EST104054.S
9	33.8	12.4	420	22	R26026	R26026 yH45c10.r1
10	33.8	12.4	308	28	AA090734	AA090734 y1540.seg
11	33.8	12.4	640	29	AA156269	AA156269 zo50c04.s
12	33.8	12.4	411	30	AA256162	AA256162 zt79b07.s
13	33.8	12.4	450	34	AA524895	AA524895 nh34g05.s
14	33.8	12.4	470	36	AA622747	AA622747 np76b05.s
15	33.8	12.4	425	41	A1026826	A1026826 ow02f12.x
16	33.8	12.4	389	42	A1125754	A1125754 qd91c01.x
17	33.8	12.4	380	43	A1200813	A1200813 qf62d03.x
18	33.8	12.4	383	43	A1214351	A1214351 q94h10.x
19	33.8	12.4	478	46	A1435410	A1435410 lh94b07.x
20	33.8	12.4	378	47	A1476629	A1476629 tm17h03.x
21	33.8	12.4	406	48	A1559724	A1559724 tq51h07.x
22	33.8	12.4	364	48	A1565082	A1565082 tq79d08.x
23	33.8	12.4	425	48	A1581345	A1581345 tq31h12.x
24	33.8	12.4	324	48	A1582930	A1582930 ts07g02.x
25	33.8	12.4	442	50	A1673758	A1673758 tw79f09.x
26	33.4	12.2	624	46	A1445609	A1445609 tj08a01.x
27	33.2	12.2	512	29	AA133301	AA133301 z117h11.s
28	33.2	12.2	451	41	AT041841	AT041841 cy34c11.x
29	33.2	12.2	474	50	AU062078	AU062078 AU062078
30	32.8	12.0	356	40	AA905453	AA905453 oj96h11.s
31	32.8	12.0	330	43	A1204595	A1204595 qd71b04.x
32	32.6	11.9	398	36	AA614641	AA614641 np54d02.s
33	32.6	11.9	424	44	A1300820	A1300820 qn47c08.x
34	32.4	11.9	547	30	AA241169	AA241169 mv24e02.r
35	32.4	11.8	457	24	N34728	N34728 yx82d10.r1
36	32.2	11.8	323	25	N98739	N98739 yy65g03.r1
37	32.2	11.8	455	25	W03383	W03383 za04b11.r1
38	32.2	11.8	452	25	W03384	W03384 za04b12.r1
39	32	11.7	408	21	T90623	T90623 yd41b01.s1
40	32	11.7	453	43	AT215594	AT215594 gh18b02.x
41	31.8	11.6	466	21	T78367	T78367 yc69g01.r1
42	31.8	11.6	544	41	A1051958	A1051958 ow83g04.x
43	31.8	11.6	322	50	A1686077	A1686077 lw98f03.x
44	31.6	11.6	469	26	W78957	W78957 zh50c03.r1
45	31.6	11.6	296	36	AA639004	AA639004 ns01e09.s

#### ALIGNMENTS

RESULT 1  
LOCUS Z84017 428 bp mRNA  
DEFINITION SSZ84017 Porcine small intestine cDNA library Sus scrofa cDNA clone  
ACCESSION Z84017  
NID 91806336  
VERSION Z84017.1 GI:1806336

KEYWORDS	EST.
SOURCE	Pig.
ORGANISM	<i>Sus scrofa</i>
	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE	Eutheria; Cetartiodactyla; Suidae; Sus.
AUTHORS	Wintero A.K., Fredholm M. and Davies W.
TITLE	(bases 1 to 428)
JOURNAL	Evaluation and characterization of a porcine small intestine cDNA
MEDLINE	library: analysis of 839 clones
COMMENT	Mamm. Genome 7 (7), 509-517 (1996) On Sep 12, 1996 this sequence version replaced gi:1404772.
FEATURES	Contact: A.K. Wintero Department of Animal Science and Animal Health, Division of Animal Genetics, The Royal Veterinary and Agricultural University Bulowsvej 13, 1870 Frederiksberg C, Denmark. Location/Qualifiers 1..428 /organism="Sus scrofa" /db_xref="taxon:9823" /clone="c12a03" /clone_lib="Porcine small intestine cDNA library" /note="directionally cloned cDNA in X1-blue MRF"
BASE COUNT	134 a        87 c        100 g        102 t        5 others
ORIGIN	
Query Match	75.5%; Score 206; DB 30; Length 428;
Best Local Similarity	83.8%; Pred. No. 8.6e-46;
Matches	227; Conservative 0; Mismatches 44; Indels 0; Gaps 0
OY	2 TTGCAATGCACCCCAATGTGCCAAGATGAACAACACTCATTCACAATAAAGACATGG 61 
DB	135 TTGCATTATGACCCCAACGCGCAGAAGATGAACAACACTCATTCAGAAATTAAGATATGG 194 
OY	62 TGACCAGGACATCTGTATCTGTGTTGAAGCTACAGAGAAAAGGATTTTATTTCAAAAAAG 121 
DB	195 TGACCAAGCACTCCATCTGTCTGTGAAGCTACAGGNAANAAGATCTTTTCCAAAAAAG 254 
OY	122 TTGCCATTGTATCTCTGAACATGGAACAACAAGGNTACATGTGACACCAAACTTG 181 
DB	255 TTGCCATTGTATCTCTGCGAAGTTGGNAGGCAAAACCGGNATRTGTGAACCAAACTTG 314 
OY	182 AGACCTACAAAATGCTGATGTTCTGTGCTGTGAGTCTANTCTCCAGGNAATGATGAC 241 
DB	315 AGACCTACAAAATGCTGATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 374 
OY	242 CCTACACTGNGCAGATGGCAACTGTGGCGA 272 
DB	375 CCTATACAGACGACAGATGGGNAACNTNGGAGA 405 
RESULT	2
AI1390317	
LOCUS	AI1390317 638 bp mRNA EST 02-FEB-1999
DEFINITION	mx03b07.y1 Soares mouse NML Mus musculus cDNA clone IMAGE:679093 5 similar to TR:O18744 O18744 LU-ECAM-1. [3] TR:O18742 TR:O18741 ; MRA sequence.
ACCESSION	AI1390317
NID	94216324
VERSION	AI1390317.1 GI:4216324
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 638) Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Peterson,B., Smaller,T., Gibbons,M., Pape,D., Harvey,N., Schuk,R., Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

**TITLE** The Mashu-NCI Mouse EST Project 1999  
**JOURNAL** Unpublished (1999)  
**COMMENT** On Aug 26, 1998 this sequence version replaced g1:3478421.

**Contact:** Marra M/Mashu-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LINDL; contact the IMAGE Consortium (infoimage.lindl.gov) for further information. This read is a RESSEQUENCE of a previously sequenced mouse clone. This read has been verified (found to hit its original self in the correct orientation)

Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gldco  
 High quality sequence stop: 461.

**FEATURES**  
 source  
 location/qualifiers  
 1..638  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="11"  
 /clone="IMAGE:679093"  
 /clone\_lib="Soares mouse NML"  
 /tissue\_type="Liver"  
 /lab\_host="DH10B"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15', TGTTACCACTGAGTGAAGGAGGCGCGCGAGATTTTTTTTTTTTTTTT 3'; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonafide."

**BASE COUNT** 180 a 167 c 145 g 144 t 2 others  
**ORIGIN**

Query Match 41.5%; Score 113.4; DB 45; Length 638;  
 Best Local Similarity 63.1%; Pred. No. 6e-21;  
 Matches 171; Conservative 0; Mismatches 100; Indels 0; Gaps 0.

QY	2	TTGCATCGACCCCATGTGCGAGAGATGAACA	CTCATTCACAAATTAAGACATGG	61
DB	154	TTGCATTAAACCCAGTGGCGGAGGAGCAAA	AGGCTCATCCAGCAATAAGAAATGG	213
QY	62	TGACCCAGGCATCTGTATCTGTTGAAGCTAC	AGGAAGCATTTATTTCAAAATG	121
DB	214	TAACTCAAGCCTCTACCTACCTGTTGAAGC	AGCCAAAGAAAGATTTCACAAACA	273
QY	122	TTGCATTTGATTCCTGAACATGSAAGACAA	AGAGTGTGAGCAACCAACTTG	181
DB	274	TAAAGATATTGTCCCGATGACCTGGAAGTCA	AAACCTGAGTAAATGCCAAAGCAG	333
QY	182	AGACCTACAAAATGCTATGTTTGGTTGGCT	ATANTCCTCAGGNAATGATGAAC	241
DB	334	AATGCTATGACAAAGCAACGTCATAGTTGG	CGAGTCCTCAGCTGCAACATGAGACGAC	393
QY	242	CCTACACTGAGCAGATGGGCAACTGTGGCGA		272
DB	394	CCTACACCCCTTCAGTATGAGACAGTGTGGGA		424

**RESULT** 3  
 AA245677 417 bp mRNA EST 10-MAR-1997  
 LOCUS mx03b07.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:679093 5'  
 DEFINITION similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.  
 ;, mRNA sequence.  
 AA245677  
 g1876466



VERSION	AA245677.1	GI:1876466
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS	1 (bases 1 to 417)	
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,	
	Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,U., Morris,M.,	
	Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,	
	Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and	
	Waterson,R.	
TITLE	The MashU-HMI Mouse EST Project	
JOURNAL	Unpublished (1996)	
COMMENT	On Apr 14, 1993 this sequence version replaced gi.693439.	
	Contact: Marra M/Mouse EST Project	
	MashU-HMI Mouse EST Project	
	Washington University School of Medicine	
	4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108	
	Tel.: 314 286 1800	
	Fax: 314 286 1810	
	Email: mouseest@watson.wustl.edu	
	This clone is available royalty-free through LNL ; contact the	
	IMAGE Consortium ( <a href="mailto:info@image.lnl.gov">info@image.lnl.gov</a> ) for further information.	
	MG1:418797	
	Seq primer: -28mJ3 rev2 ET from Amersham	
	High quality sequence stop: 399.	
FEATURES	Location/Qualifiers	
SOURCE	1..417	
	/organism="Mus musculus"	
	/db_xref="taxon:10090"	
	/map="12"	
	/clone="IMAGE:679093"	
	/clone_lib="Soares mouse NML"	
	/tissue_type="liver"	
	/lab_host="DH10B"	
	/note="Vector: p17T3D-Pac (Pharmacia) with a modified	
	polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA	
	was primed with a Not I - oligo(dT) primer [5'	
	TGTTACCAATCTGAAGTGCGGCAGCCGGCAGATTCTTTTTTTTTTTTTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors	
	(Pharmacia), digested with Not I and cloned into the Not I	
	and Eco RI sites of the modified pT73 vector. Library	
	constructed and normalized by Bento Soares and M.Fatima	
	Bonaldo."	
BASE COUNT	118 a 112 c 94 g 93 t	
ORIGIN		
Query Match	36.5%; Score 99.6; DB 30; Length 417;	
Best Local Similarity	62.8%; Pred. No. 3e-17;	
Matches 167; Conservative 0; Mismatches 99; Indels 1; Gaps 1;		
OY 2 TTGCATCGACCCCACATGTGCCAGAATGAAAACACTCATTCACAATAAAGACATGG 61		
Db 153 TTGCATTAACCCCCAGTGTGCC-GGAGACGAAGGCTCATCCACAGCAAAGAAATGG 211		
OY 62 TGACCAGCACTCTCTGTATCTGTTTGAAGCTACAGGAAGAAAGCATTTTATTCAAAAATG 121		
Db 212 TTAACCAAGCCCTACACCTACCTGTTTGAAGCCAGCCAAGGAAGAGTTTATTTTCAAAAAA 271		
OY 122 TTGCATTTTGAATTCCTAAACATGGAAGACAAAGAGNTGACTATGTGAGAACCAAACTTG 181		
Db 272 TAAGCATTTTAGTCTCCGANTGACCTGGAGTCAAAACCTGAGTACTTAATGCCAAAAACGAG 331		
OY 182 AGACCTACAAAATCTGATGTTGTGTTGCTGAGCTTANTCTTCAGGNAATGATGAAC 241		
Db 332 AATCTATGACAAACACACGTCATAGTTGCGAGATCTTCACCTGCAACATGGAGACGACC 391		
OY 242 CCTACACTGNCAGATGGGCAACTGT 267		
Db 392 CCTACACCTTCAGTATGACAGTGT 417		

RESULT	4	109430/c	109430	441 bp	mRNA	EST	03-AUG-1993
LOCUS	109430	441 bp	mRNA	EST	03-AUG-1993		
DEFINITION	ESR07c723	Infant Brain, Bento Soares	Homo sapiens	CDNA clone	HIBBU20		
ACCESSION	5'	end, mRNA sequence.					
NID	109430						
VERSION	9390458						
KEYWORDS	109430.1	GI:390458					
SOURCE	EST.						
ORGANISM	human.						
REFERENCE	Homo sapiens						
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;						
TITLE	Eutheria; Primates; Catarrhini; Homidae; Homo.						
JOURNAL	1 (bases 1 to 441)						
MEDLINE	Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.						
COMMENT	Rapid cDNA sequencing (expressed sequence tags) from a						
	directionally cloned human infant brain cDNA library						
	Nature Genet. 4, 373-380 (1993)						
	94004965						
FEATURES	Source						
	Contact: Adams, MD						
	The Institute for Genomic Research						
	932 Clopper Road, Gaithersburg, MD 20878						
	Tel: 3018699056						
	Fax: 3018699423						
	Email: mdadams@tigr.org						
	Seq primer: M13 Reverse.						
	Location/Qualifiers						
	1..441						
	/organism="Homo sapiens"						
	/db_xref="ATCC (lnhost):85877"						
	/db_xref="taxon:9606"						
	/clone="HIBBU20"						
	/clone_lib="Infant Brain, Bento Soares"						
	BASE COUNT 108 a 92 c 114 g 125 t 2 others						
	ORIGIN						
	Query Match 13.2%; Score 36; DB 20; Length 441;						
	Best Local Similarity 69.6%; Pred. No. 3.5;						
	Matches 48; Conservative 0; Mismatches 21; Indels 0; Gaps 0;						
OY	51	AAAGACATGCTGACCCAGCATCTCTATCTGTTTAAGACTACAGAAAGCATTTTA	110				
DB	72	AAAGACATGCTGACCCAGCATCTCTATCTGTTTAAGACTACAGAAAGCATTTTA	13				
OY	111	TTTCAAAA 119					
DB	12	TTTCAAAA 4					
RESULT	5						
LOCUS	AA645367	265 bp	mRNA	EST	28-OCT-1997		
DEFINITION	VS80905.f1	Breast mouse myotubes MRLB5	Mus musculus	CDNA clone			
ACCESSION	IMAGE:1152632	5' similar to:GI066285	GI066285	HYPOTHETICAL	15.2		
NID	AA645367						
VERSION	92571796						
KEYWORDS	AA645367.1	GI:2571796					
SOURCE	EST.						
ORGANISM	house mouse.						
	Mus musculus						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;						
	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
	1 (bases 1 to 265)						
	Matta,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,						
	Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,						
	Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,						
	Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and						
	Waterson,R.						





```

/organism="Homo sapiens"
/db_xref="GB:538321"
/db_xref="taxon:9606"
/clone="IMAGE:132690"
/clone_lib="Soares placenta NB2HP"
/sex="female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Placenta; Vector: pT73D (pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) 3', AACGGAGAGAAATTCGCGCCGACAGAAATTTTCTTTTCTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73D vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      112 a      70 c      103 g      129 t      6 others
ORIGIN

Query Match      12.4%; Score 33.8; DB 22; Length 420;
Best Local Similarity 51.0%; Pred. No. 14;
Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY  108 TTATTTCAAAAATGTTGCCATTTTGTCTTCCTGAACATGGAAGACAAAGGATGATGT 167
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  202 TTAATTTTAAAGCAGCTAGCAAAATAAATGACTCATTTATCAATCAAGTCCAAATTA 143
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  168 GAGACCAAAACTTGAGACCTACACAAAATCCTGATGTTGTGGTTCGAGCTANTCCTCC 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  142 AAAACCTAAATAATGGTAATATATATATTAAGCTGAGTAATATGTGGCTGTGATAAGCCCATG 83
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  228 AGGNAATGATGAACCTACACTGNG 252
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  82 AGGTAATATGAACACTAAACTTGG 58
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT  10
LOCUS   AA090734.c
DEFINITION  AA090734 308 bp mRNA
          y1540.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
          5', mRNA sequence.
ACCESSION  AA090734
NID       91635310
VERSION   AA090734.1 GI:1635310
KEYWORDS  EST.
SOURCE    human.
          ORGANISM
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 308)
          TITLE       Liew,C.C.
          AUTHORS     Liaw,C.C.
          JOURNAL     cDNAs from fetal heart (1996)
          COMMENT     On Sep 12, 1996 this sequence version replaced gi:1400802.

Contact: Liew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Bantling Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewcc@utcc.utoronto.ca

PCR Primers
FORWARD: 5' GCGAAGCTGGAATTAACCCCTCACTAAAGG 3'
BACKWARD: 5' CCACTGAATGTATTAAGCACTCAATGAGCG 3'
Seq primer: 5' GAATTAATACCCCTCACTAAAGG 3'.
Location/Qualifiers
1..308
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human fetal heart, Lambda ZAP Express"

```

Query Match	12.48	Score 33.8	DB 28	Length 308
Best Local Similarity	51.0%	Pred. No.14		
Matches 74; Conservative	0	Mismatches 71	Indels 0	Gaps 0
Db	108	TTATTCAAAATGTTCCATTTTGATTCCTGTAACATGGAACACAAAGAGTACTATGT	167	
Db	132	TTATTTTAAAGCACTTGAATATAATGACTGATTTATCATCCAAAGTGTCAAACATTA	93	
Db	168	GAGACCAAACTTGAGACTACAAAATGCTGTTGTTGCTGAGCTTANTCTCC	227	
Db	92	AAAACCTAAAAATGTAATCATATATTAAGCTGAGTAATATGCGCTGATTAAGCCCATG	33	
Db	228	AGGNAATGATGAAACCCACACCTGNG	252	
Db	32	AGGTAAATATGAAACCTAAACTTGG	8	
RESULT 11	AA156269	640 bp	MRNA	EST 11-DEC-1996
LOCUS	AA156269/c	2050c04.s1	stratagene endothelial cell 937223	Homo sapiens CDNA
DEFINITION	AA156269	AA156269	clone IMAGE:590310	3', mRNA sequence.
ACCESSION	NID	91727904	AA156269.1	GI:1727904
VERSION	KEYWORDS	EST.		
SOURCE	ORGANISM	human.		
REFERENCE	AUTHORS	Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
		Eutheria; Primates; Catarrhini; Homiidae; Homo.		
		1 (bases 1 to 640)		
		Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,		
		Chisoe,S., Dietrich,N., Dubuque,T., Favellio,A., Gish,W.,		
		Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,		
		Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,		
		Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thilly,Meg,J.,		
		Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.		
		and Marra,M.		
TITLE	JOURNAL	Generation and analysis of 280,000 human expressed sequence tags		
MEDLINE	97044478	Genome Res. 6 (9), 807-828 (1996)		
COMMENT		On Sep 12, 1996 this sequence version replaced gi:1397818.		
FEATURES	SOURCE	Contact: Wilson RK		
		Washington University School of Medicine		
		4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
		Tel: 314 286 1800		
		Fax: 314 286 1810		
		Email: estewatson.wustl.edu		
		This clone is available royalty-free through LNL; contact the		
		IMAGE Consortium (info@image.llnl.gov) for further information.		
		Seq primer: -40M13 fwd. from Amersham		
		High quality sequence stop: 312.		
		Location/Qualifiers		
		1..640		
		/organism="Homo sapiens"		
		/db_xref="GDB:4621696"		
		/db_xref="taxon:9606"		
		/map="7: 17621"		
		/clone="IMAGE:590310"		
		/clone_lib="Stratagene endothelial cell 937223"		
		/dev_stage="umbilical vein, 1 passage"		

```

./lab_host="SOLR (kanamycin resistant)"
/Note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt.
Unidirectional vein endothelial cells, passaged once. Average
Insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor
sequence: 5' GAATTCGGACGAG 3' -3' adaptor sequence: 5'
CTCGAGCTTTTCTTTTCTTTTCTTTT 3'"

```

```

BASE COUNT      139 a      139 c      175 g      181 t      6 others
ORIGIN

```

```

Query Match      12.4%; Score 33.8; DB 29; Length 640;
Best Local Similarity 68.1%; Pred. No. 14;
Matches 47; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```

```

QY  51 AAAGCAGTGTGACCGACATCTGTATCTGTTGAAGTACAGAAAGCATTTTA 110
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   81 AAAGCAGAGTGAACAGATATCGCTTACCTGTGAACATAAATAAGCTTGTTA 22
QY  111 TTTCAGAAA 119
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   21 TTTCAGAAA 13

```

```

RESULT 12
LOCUS   AA256162      411 bp      mRNA      EST      06-AUG-1997
DEFINITION  z179b07.s1 Soares.NhMpu_S1 Homo sapiens cDNA clone IMAGE:681877
ACCESSION  AA256162
VERSION    91891701
KEYWORDS   AA256162.1 GI:1891701
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 411)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S.,
           Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
           Moore,B., Scheilenberg,R., Steptoe,M., Tan,F., Theisling,B.,
           White,Y., Wylie,T., Waterston,R. and Wilson,R.
           MASHU-Merck EST Project 1997
           Unpublished (1997)
           On Sep 12, 1996 this sequence version replaced gi:1288125.
COMMENT
JOURNAL
TITLE
JOURNAL
COMMENT

```

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 602 Std Error: 0.00
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 357.
Location/Qualifiers
1. 411
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="14"
/clone="IMAGE:681877"
/clone_lib="Soares_NhMpu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/Note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA

```

```

was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

```

```

BASE COUNT      104 a      89 c      106 g      112 t
ORIGIN

```

```

Query Match      12.4%; Score 33.8; DB 30; Length 411;
Best Local Similarity 68.1%; Pred. No. 14;
Matches 47; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```

```

QY  51 AAAGCAGTGTGACCGACATCTGTATCTGTTGAAGTACAGAAAGCATTTTA 110
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   72 AAAGCAGAGTGAACAGATATCGCTTACCTGTGAACATAAATAAGCTTGTTA 13
QY  111 TTTCAGAAA 119
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   12 TTTCAGAAA 4

```

```

RESULT 13
LOCUS   AA524895      450 bp      mRNA      EST      05-AUG-1997
DEFINITION  nh34g05.s1 NCI_CGAP_P3 Homo sapiens cDNA clone IMAGE:954296, mRNA
ACCESSION  AA524895
VERSION    92265823
KEYWORDS   AA524895.1 GI:2265823
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 450)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
           National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
           Unpublished (1997)
           On May 5, 1995 this sequence version replaced gi:798270.
COMMENT
JOURNAL
TITLE
JOURNAL
COMMENT

```

```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/dbcp/image/image.html
Insert Length: 513 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 428.
Location/Qualifiers
1. 450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone="IMAGE:954296"
/clone_lib="NCI_CGAP_P3"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/Note="Vector: PAMPIO, Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected cells
histologically-determined to be fully malignant prostate
cancer cells. Double-stranded cDNA was ligated to EcoRI

```





**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 21:05:17 : Search time 3264.21 Seconds  
(without alignments)  
217.268 Million cell updates/sec

Title: US-09-049-696-1  
Perfect score: 223  
Sequence: 1 GAATCAGCAGGAGATGTAC.....ATCTCTGATCTGTTTGAG 223

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_dal:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pil:\*  
8: gb\_p12:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hgt:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hgt1:\*  
35: gb\_hgt2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	221.4	99.3	3311	11	AF039400	Homo sapi

2	203	91.0	2826	42	AF127036	AF127036 Homo sapi
3	185.4	83.1	35278	11	AF039401	AF039401 Homo sapi
4	138.2	62.0	2937	12	AB017156	AB017156 Mus muscu
5	91	40.8	3317	3	AF001261	AF001261 Bos tauru
6	91	40.8	2820	3	AF001263	AF001263 Bos tauru
7	91	40.8	1265	3	AF001264	AF001264 Bos tauru
8	89.4	40.1	2984	3	BFU36445	BFU36445 Bos tauru
9	86.6	38.8	3288	3	AF001262	AF001262 Bos tauru
10	85.2	38.2	3415	42	AF03976	AF03976 Homo sapi
11	72.2	32.4	3022	12	AF047838	AF047838 Mus muscu
12	72.2	32.4	3471	12	AF052746	AF052746 Mus muscu
13	49.2	22.1	3604	9	AB026833	AB026833 Homo sapi
14	40.6	18.2	330742	17	PBU42580	U42580 Paramecium
15	35.4	15.9	135039	11	AC006060	AC006060 Homo sapi
16	34	15.2	1551	3	BTAT	X35604 B. taurus mR
17	33.6	15.1	5677	9	AB020629	AB020629 Homo sapi
18	33.4	15.0	1674	12	MSM222714	AJ222714 Mus muscu
19	33.4	15.0	28396	36	CELF10E7	U41264 Caenorhabdi
20	33.2	14.9	1848	12	MUSNARS	L37663 Mus musculu
21	33	14.8	6924	5	135776	135776 Sequence 2
22	33	14.8	7941	9	HUMPTPRZ	M33426 Human prote
23	33	14.8	1839	10	HSPTPZ	X54135 Human HPTP
24	33	14.8	7941	14	G20044	G20044 SMS51987 Er
25	32.6	14.6	1207	8	AF139448	AF139448 Gossypium
26	32.6	14.6	151770	10	HSAC002066	AC002066 Human BAC
27	32.6	14.6	2106	12	RATNARAD	I31619 Rattus ratt
28	32.6	14.6	3030	12	S53887	S53887 nicotinic r
29	32.6	14.6	424	14	G40980	G40980 LSCV22 Capr
30	32.6	14.6	424	14	G41003	G41003 LSCV37 Capr
31	32	14.3	137947	31	HS4514	AL023581 Human DNA
32	32	14.3	148133	34	HS667H12	AL035414 Homo sapi
33	31.8	14.3	1624	1	PMAROA	Z14100 P. multocida
34	31.8	14.3	4438	1	STOAPCE	D63788 Synechococc
35	31.8	14.3	1201	8	AF139449	AF139449 Gossypium
36	31.8	14.3	1201	8	AF139450	AF139450 Gossypium
37	31.8	14.3	176081	35	AC006983	AC006983 Homo sapi
38	31.8	14.3	35840	37	CELC27H5	U14635 Caenorhabdi
39	31.6	14.2	55090	9	AP000148	AP000148 Homo sapi
40	31.6	14.2	32196	9	AP000235	AP000235 Homo sapi
41	31.6	14.2	190653	35	AC006994	AC006994 Homo sapi
42	31.4	14.1	4804	9	AB023175	AB023175 Homo sapi
43	31.4	14.1	1559	9	HUMATNAR	L25827 Human a7 ni
44	31.4	14.1	2087	10	HSARA7A	X70297 H.sapiens m
45	31.4	14.1	1555	10	HSCRNA7A	Z23141 H.sapiens c

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS		AF039400				
DEFINITION		Homo sapiens calcium-dependent chloride channel-1 (hclcal) mRNA, complete cds.				
ACCESSION		AF039400				
NTD		94009457				
VERSION		AF039400.1	GI:4009457			
KEYWORDS						
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,R.D., Fuller,C.M. and Pauli,B.U.				
TITLE		Genomic cloning, molecular characterization, and functional analysis of human ClCa1, the first human member of the family of Ca2+-activated Cl- channel proteins				
JOURNAL		Genomics 54 (2), 200-214 (1998)				
MEDLINE		99047526				
REFERENCE		2 (bases 1 to 3311)				
AUTHORS		Gruber,A.D., Elble,R. and Pauli,B.U.				
TITLE		Direct Submission				

JOURNML Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers  
source

1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hclcal"  
352. .3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPKSSVFILIHLEGLALSNSLIQUNNGYEGIVVADPNP  
EDTLIOIKDMVTOASLYFEATGKRFENVAIILPETWKIRADYVRKLETYNA  
DVLVAESTPPENDEPYTEOMGNGCEKGERHILPDLFAGKRLAEGPOGRAEYEMAH  
LRMGVDEYNDEKFTLSNGRIQAVRCSAGITGNVVKCGGSCYTRKCTFNVGL  
YKGECEFLVSRQTEKASIMFAQVDSIVECTEONHNKEAPNOKNOKNLSWEVI  
RDSDEFKTPPTOPNPPTFSLIQGRIVCYLDKSGSMATGNRLNRLNOGQFL  
LQVEIGSWGWTDFSAHVSQELIQLNSGSDRLAKRLPAASGCTISGILBSA  
FTVIRKRYPTDGSIEIVLLTDEGNTISGCRFVQSAIITHVALGPSAAOELELSK  
MTGGLQTVASDVQVONGLIDAFGLSSGNGAVORSIQLESGLTLQNSQMNMTYIV  
DSTVGKDLFLITWTOPPOLILMDPSGOGGQFVVDKNTKMAVLIQIPGIAKVGTVK  
SLQASQTLITVTSRASNAITLPITVTSKNTDTSKPSPLVYANIRGASPIILRA  
SVTALIESVNGKVTLELDNGAGADATKDGVSRYETTYDNGRYSVYRVLGYN  
AARRVITPOOSGALITPGWIENDEIOMNPPREINKDVOHKVCSRSISGSEFVAS  
DVPNAPITPDLFPPGQITDLKAEIHGSLINLTWTAPEGDDYDHTAKYIIRISTSLD  
LRDFNESLOVNTTALIPKEANSEVLEFPENITFENGIDLFIAIOAVDKVLKSEI  
SNARVSLFIPQTPPETPSPDETSAFCPIHINSTIPGIHLIKMKWIGELQSLIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 99.3%; Score 221.4; DB 11; Length 3311;  
Best Local Similarity 99.6%; Pred. No. 4e-56;  
Matches 222; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAATCAGAGGAGATGTACAGCAATGGGCCATTAGAGTTCTGTTCACTTCAAT 60  
DB 328 GGAATCAGAGGAGATGTACAGCAATGGGCCATTAGAGTTCTGTTCACTTCAAT 387  
QY 61 CTCACCTTTAGAAAGGGCCCTGAGTAATTCATTCATTCAGTGAACAAATGGCAT 120  
DB 388 CTCACCTTTAGAAAGGGCCCTGAGTAATTCATTCATTCAGTGAACAAATGGCAT 447  
QY 121 GAAGGCAATGTGCTTGAATGACCCCAATGTGCCAAGAAGTGAACAACCTATTCACAA 180  
DB 448 GAAGGCAATGTGCTTGAATGACCCCAATGTGCCAAGAAGTGAACAACCTATTCACAA 507  
QY 181 ATAAAGGCAATGTGACCCAGGCAATCTGTATCTGTTGAAG 223  
DB 508 ATAAAGGCAATGTGACCCAGGCAATCTGTATCTGTTGAAG 550

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CaCC1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
AUTHORS Agnel,M., Vermet,T. and Culouscou,J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel

JOURNML Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel,M. and Culouscou,J.-M.  
TITLE Direct Submission  
JOURNML Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Neuilly-Malmaison 92500, France

FEATURES  
source  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="CaCC1"  
5. .2749  
/note="CaCC1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AA025487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPKSSVFILIHLEGLALSNSLIQUNNGYEGIVVADPNP  
EDTLIOIKDMVTOASLYFEATGKRFENVAIILPETWKIRADYVRKLETYNA  
DVLVAESTPPENDEPYTEOMGNGCEKGERHILPDLFAGKRLAEGPOGRAEYEMAH  
LRMGVDEYNDEKFTLSNGRIQAVRCSAGITGNVVKCGGSCYTRKCTFNVGL  
YKGECEFLVSRQTEKASIMFAQVDSIVECTEONHNKEAPNOKNOKNLSWEVI  
RDSDEFKTPPTOPNPPTFSLIQGRIVCYLDKSGSMATGNRLNRLNOGQFL  
LQVEIGSWGWTDFSAHVSQELIQLNSGSDRLAKRLPAASGCTISGILBSA  
FTVIRKRYPTDGSIEIVLLTDEGNTISGCRFVQSAIITHVALGPSAAOELELSK  
MTGGLQTVASDVQVONGLIDAFGLSSGNGAVORSIQLESGLTLQNSQMNMTYIV  
DSTVGKDLFLITWTOPPOLILMDPSGOGGQFVVDKNTKMAVLIQIPGIAKVGTVK  
SLQASQTLITVTSRASNAITLPITVTSKNTDTSKPSPLVYANIRGASPIILRA  
SVTALIESVNGKVTLELDNGAGADATKDGVSRYETTYDNGRYSVYRVLGYN  
AARRVITPOOSGALITPGWIENDEIOMNPPREINKDVOHKVCSRSISGSEFVAS  
DVPNAPITPDLFPPGQITDLKAEIHGSLINLTWTAPEGDDYDHTAKYIIRISTSLD  
LRDFNESLOVNTTALIPKEANSEVLEFPENITFENGIDLFIAIOAVDKVLKSEI  
SNARVSLFIPQTPPETPSPDETSAFCPIHINSTIPGIHLIKMKWIGELQSLIA  
"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 91.0%; Score 203; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 1.2e-50;  
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 AGCAATGGGCGCATTTAGAGTTCTGTTCACTTCACTTCTAGAGGGGC 80  
DB 1 AGCAATGGGCGCATTTAGAGTTCTGTTCACTTCACTTCTAGAGGGGC 60  
QY 81 CCTAGTAATTCATCTTACGTCGACCAACAATGCGTATGAAGGCAAT 140  
DB 61 CCTAGTAATTCATCTTACGTCGACCAACAATGCGTATGAAGGCAAT 120  
QY 141 CGACCCCAATGTGCGCAGAGATGAACACTCATCAACAATAAGAGCATGGTACCCA 200  
DB 121 CGACCCCAATGTGCGCAGAGATGAACACTCATCAACAATAAGAGCATGGTACCCA 180  
QY 201 GGCATCTCTGTATCTGTTGAAG 223  
DB 181 GGCATCTCTGTATCTGTTGAAG 203

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hclcal) gene,  
DEFINITION complete cds.

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	source
AF039401.1	GI:4009459	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 35278)	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuz,R.D., Fuller,C.M. and Pauli,B.U.	Genomic cloning, molecular characterization, and functional analysis of human CCLAI, the first human member of the family of Ca2+-activated Cl- channel proteins	Genomics 54 (2), 200-214 (1998)	2 (bases 1 to 35278)	Gruber,A.D., Elble,R. and Pauli,B.U.	Direct Submission	Submitted (13-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA	Location/Qualifiers	1. .35278
														/organism="Homo sapiens"
														/db_xref="taxon:9606"
														/chromosome="1"
														/map="1p22-p31"
														1. .1617
														/gene="hCCLAI"
														1. .33522
														/gene="hCCLAI"
														767. .1265
														/rpt_family="11 transposable element"
														/rpt_type="dispersed
														1587. .1596
														/gene="hCCLAI"
														/note="putative"
														1618. .1708
														/gene="hCCLAI"
														join(number=1
														join(1618. .1708,1962. .2383,6667. .6807,6979. .7126,9696. .9801,15454. .15631,18591. .18809,19783. .20010,22253. .22427,24522. .24628,26644. .26859,27447. .27708,28766. .28936,28766. .28936,31840. .32079,32919. .33522)
														/gene="hCCLAI"
														/product="calcium-dependent chloride channel-1"
														1962. .2383
														/gene="hCCLAI"
														/number=2
														join(2222. .2383,6667. .6807,6979. .7126,9696. .9801,15454. .15631,18591. .18809,19783. .20010,22253. .22427,24522. .24628,26644. .26859,27447. .27708,28766. .28936,31840. .32079,32919. .33310)
														/gene="hCCLAI"
														/codon_start=1
														/product="calcium-dependent chloride channel-1"
														/protein_id="AAC95429.1"
														/db_xref="PID:94009460"
														/db_xref="GI:4009460"
														/translation="MGPKRSVFLILHLLEGALSNLSIQLNNGYECIVADPNP/EDFLIQIKDMVTQASLYLEATGKRREYNVALIPETWTKTADYVRPYLEYKNAALDVLAESIPENDEPYTEOMGNCGEKGRIRILTPDFTAGKLABYGGKAFVEMHMLRMVAFDEYNDEKEFYLSNGRIQAVRCAGAGITGNVYKCGGSGCYRCGFNVETGLYKCEPFLIOREREKASIMPAOHVDSIVERCTEONHKEAPKNONOKNLRSTWEIYLRDSEDKTKTPMTQPPRPTSLQIQIRIYCLDKSGMATCNRLNRLNQAQLTLQYLELGSWGAVTFDSAAHVQSELIQINSGSDHDTLAKRIPAAASGDTSCSLRSVAFVIRKRPIDGSEIVLTDGEDNTISCFENFEVQSGAILHTVALGSAQAELELSKFTGGLQYASDOVONNGILIDAFGALSSNGNVSORSQILESGKGLTNSOMMNTVLSIDSYGKDTLFLITWTQPOPLLMDPSOKGCGVYDKNTMAQLQIPGILKVTAKKSLQASQTLFLITVSRASNALPLPYTASNKDPSKRPSPLYVYANRQASPLILAAKSVTLIIEVNGKTYTIELNDGACADKDCVSRFTYDNGRYSVYRALGCVNAAARRVYIDQSGGALITPEWINDETQWNPPELNKDDVQKQVCFRTSSGGSFVASDVNPAPLPDLPFGQITDLNAEIHGSGSLINTWAPAGDDYDHGTANHYYLIRISITLID

Query Match	Best Local Similarity	83.1%	Score 185.4	DB 11	Length 35278
Matches 186	Conservative	0	Mismatches 1	Indels 0	Gaps 0
Qy 1	GAATCACAGGAGATGTACAGCAATGGGCGCATTTAGAGTTCTGTGTCATCTTGATT	60			
Db 2198	GGAATCACAGGAGAGATGTACAGCAATGGGCGCATTTAGAGTTCTGTGTCATCTTGATT	2257			
Qy 61	CTTCACTTTAGAAAGGGGCCCTGAGTATTCATCTCATCTGACCTGACACAAATGGCTAT	120			
Db 2258	CTTCACTTTAGAAAGGGGCCCTGAGTATTCATCTCATCTGACCTGACACAAATGGCTAT	2317			
Qy 121	GAAGGATGTCGTGCATGACACCCCAATGGCCAGAAAGTGAACACCTCATTCACAA	180			
Db 2318	GAAGGATGTCGTGCATGACACCCCAATGGCCAGAAAGTGAACACCTCATTCACAA	2377			
Qy 181	ATTAAGG 187				
Db 2378	ATTAAGG 2384				

BASE COUNT 11245 a 6870 c 7230 g 9933 t

ORIGIN

LRDKNSELPQNTALIPKRNSEVFLFPENITTFENGDTLFAIAVDVDDKSEI  
 SNIRVSLFIPQPTPEPSPDEISACPNHINSLTIPGHIHKIMRWIGEDLDSIA

LOCUS AB017156 2937 bp mRNA ROD 04-MAR-1999

DEFINITION Mus musculus gob-5 mRNA, complete cds.

ACCESSION AB017156

NTD 93721911

VERSION AB017156.1 GI:3721911  
KEYWORDS GOB-5.  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Mammalia; Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (sites)  
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice  
JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
MEDLINE 99160866  
REFERENCE 2 (bases 1 to 2937)  
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) to the DBJ/EMBL/GenBank databases. Tohru Komiyama, Eikuro, JST, Hirohashi Cell Configuration Project; 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan (E-mail:tkomihccp.jst.go.jp, Tel:81-298-47-7563, Fax:81-298-47-5226)  
FEATURES  
SOURCE Location/Qualifiers  
1..2937  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15..2756  
/gene="gob-5"  
15..2756  
/gene="gob-5"  
/codon\_start=1  
/protein\_id="BA03743.1"  
/db\_xref="PID:g1034712"  
/db\_xref="PID:g3721912"  
/db\_xref="GI:3721912"  
/translation="MESKSPVLLILHLGVLSESLIOLNNNGEIVATIDHVP  
EDALIQIKIDVTQASPYLEPATKRRYEFKNVAILIPESKAPETIRPLLEFKNA  
DVLVSTSPGLNDEPYTEHIGACGKGRIRHLTPDFLAGKKLTQGPDRDFVEMAH  
FRMGVFNENDEKYLKRGKAVRCASAIATKNQVRCOGSCITNGKCVIDRWG  
LYKNCVFPDPHONKASIMFNQVNSVECTEKNHNOAEPNDONRCMLRSTWEG  
LOESPEKOTPMFACPPAPTEFSLIIGORIVCLVDKSGSLNDRRLNRNQA  
LIQTYECSWGMVTFDSAAVYOSLKGNSGADRLIKHLPTYSAGSTICGLRT  
AFVTKKTPIDGSEVILTBDEDTISCDLVYQSAIITHVALGSRRAKLEPOLS  
KMTGLQYSSDQVONGLVDAFALSSGNAIAQHSIQLESRGVNLNOMNGSVI  
VDSVGRDLEFLITWTHTPTLFIIMDPGVDQNGFLIDTTKVALQVGTAKGEMK  
YSIQASQTLITVTSRAASATLPIITVTPVKNKTKGFPSPVIVASIRGASPLR  
ASVATLIESVNGKVTLELDNGACADATKNDGYSRFEAFDANGRSYRIMALGV  
TSDRRAAPKRRAMVTDGWIJEDGVRNPPRPENSYODKOLCFRTSSGSEVATN  
VFAAPFIDLPFPCCITDKASIOGONLVNTLTPAPGDDYDHGRASNYITMSTIVD  
LNDHNTSLQVNTGLIPREASSEIFEFELGNTFGKGTDFITIAQVXKSNLSEI  
SNIRVSVFIPQEPEDSTPPPCPDISTINSTIPGIVHLKIMRWLEMOVTTLSLH

Db 195 TCTCCATCCTGTTGAAG 213  
RESULT 5  
AF001261  
LOCUS AF001261  
DEFINITION Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
complete cds.  
ACCESSION AF001261  
NID 92623762  
VERSION AF001261.1 GI:2623762  
KEYWORDS Bos taurus.  
SOURCE Bos taurus.  
ORGANISM Bos taurus; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eukaryota; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 3317)  
AUTHORS Eble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3317)  
AUTHORS Eble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA  
FEATURES  
SOURCE Location/Qualifiers  
1..3317  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="1"  
/tissue\_type="lung"  
63..2780  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86529.1"  
/db\_xref="PID:g2623763"  
/db\_xref="GI:2623763"  
/translation="MVICLVILFLHLPGKSSVNLINNGYDGIIVAINPSPE  
DEKLIRKMTVEASTYLPATKRRYFENVSLIIPMKSESEIFPQESYDAD  
VIAVPLKIGDPTITQYRCGKGIHTFTNPLLNPHYSGRGVFEVEMAH  
RWGIFEDVNDQPIYSRNTIATRSTHTIGINVFRCPEGSCITSLCRDSOTG  
LYEAKCFELPKSOTAKESIMFPLSLHSVFCTETHTNTEANLQNMKNGSTVDY  
IMNSVQONTSPMTENMPYHPFSLSKQRYVAVCLDKSGMSAEDEKLFOMQAAE  
LYILOVEKESLNGVTFEDSAEIONHLTITDNYOKITAKLPQVANGSTICGL  
KAGQAIIHSDQSGSEIILLDGEENINSCPEDEVKRGKAIHTALPQASAKELE  
TLSNMGIRYFPAKNDITGTGNAFSRSSSGSITQOALQESKALITGRKRVNGV  
PVDSVGNDFEYVATWTKPEYVLDQPKKRTISPRDKNLIRBARQIPIGNET  
GTWYTSLLNNHASSQMLTVTTTARSPTLIPVATLHMSQHTAHFSPHIVAYQVSQ  
GFLEPLGISVIAIETEDGHQVLELDNNGAGRTVYNDGIYSRTFDYDNGRTSLK  
VHAQARNNTARLNLROPNKRVLYVGVENGKIIINPREVVDLAKAIDEFSRLT  
SGGSFTYSGAPPEGNHSPSPSKITDLEAKFKEDYIOLSMTAGVLDKANSYII  
RISFPMRDGPDGNHNTLVNTSNLIPREASKEFEKPEHPFVRENTKFIYVOAIN  
EANLISEVSHVOAIKFIPLPEDSVHLGKIRISITLAILGLMIPSVF"

Db 103 ATCTCTGCTGG---AATGAAAAGTTCAATGGTAATTTGATTACATGGGTATCATG 159  
QY 125 GCATGTGCGTGAATGACGACCCCATGTCGACAGATGAAGACATTCATCAACAATAA 184  
Db 160 GCATGTGCGTGAATGACGACCCCATGTCGACAGATGAAGACATTCATCAACAATAA 219  
QY 185 AGGACATGTGACCCAGCATCTCTGATCTGTTGAAG 223  
Db 220 AGGAAATGTAAGTGAAGCTTCTACTTACCTGTTTCATG 258

RESULT 6  
AF001263 2820 bp mRNA MAM 19-NOV-1997  
LOCUS Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.  
ACCESSION AF001263  
NID 92623766  
VERSION AF001263.1 GI:2623766  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 2820)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
JOURNAL endothelial chloride channel  
AUTHORS Unpublished  
TITLE 2 (bases 1 to 2820)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA

FEATURES  
source Location/Qualifiers  
1. .2820  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone="3"  
/tissue\_type="lung"  
/cell\_type="aortic endothelial cells"  
194. .2656  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86531.1"  
/db\_xref="PID:92623767"  
/db\_xref="GI:2623767"  
/translation="MVICLVNLIFFLHLPGKSSMNLINNGYDGIYAIINPSVE  
DKLEIKEMYTEASTLYFHATKRVRVFNYSILIPMTWKSSTYFTPKDSYQAD  
VIVANPYLYGDDPTLYGRCGEKRTIHTPNLITNNPHYSGRGVPTHEVAHL  
RWGIDETVNDQPFYISRNKTIETATCSHTIGINVFYKPCPGSCITSLCRDSOTG  
LYEACTFLPKRSQAKESIMPELSHSTVECTEKTHTNTEAPNLONKSNKSTWDY  
INNSVDFOPTSPTEMNPTHTPELSKQRVCLVDKSGMSAEDLFPOMQAE  
KLVIOYIEKSGLVGWTVDVAEIOHNLRTIDDNVYOKITAKIPQVANGSTISGL  
KGFQALIHSDSTSGSEIILITDGEDNINSGFEVYKRSKALIHITIALGSPAARLE  
TISNMGGRYFRANDITGLTNAFSRISRSSGSIITQAOLESKAKITGRKRVNTY  
PVDSTVGNDTFVVTWITOKPEIIVDDPRGKRTSDREFEDKLNTISAKLIPLIET  
GTWYSILNNHASSQMLIVTVTRAPSPITPIVIAAHMSOHTAHPSPMIYAAVSQ  
GFLPVGLISVIAIIEIEDGHQVLELMDAGADIVKNDGISRYFTDYDYGNGRSLK  
VVAQAKNTARLNIROPKNVLYVPGYVNGKTIILPPREVKADLAKAKIDFESLT  
SGSFTVSGAPPGNHPSPSKITDLAKFKEDYIQIOWTAHPGVLDKXKASGSP  
MRFSHQVAKVLELOQHSFO"

BASE COUNT 940 a 365 c 360 g 755 t  
ORIGIN

Query Match 40.8%; Score 91; DB 3; Length 2820;  
Best Local Similarity 66.7%; Pred. No. 3e-17;

Matches 146; Conservative 0; Mismatches 70; Indels 3; Gaps 1;  
QY 5 TCACAGGAGATGTACAGCAGATGGGCCATTAGAGCTTGTCATTCGATTCCTTC 64  
Db 174 TTAGTACATGTGCAAAAATGTGCTGCTGTAATGTTATCTGTTCTTAACCTTC 233  
QY 65 ACCCTTAGAAGGGGCCGAGTAATCACTCATTCAGCTGAACACANTGGCTATGAG 124  
Db 234 ATCTCTGCTGG---AATGAAAAGTTCAATGGTAATTTGATTACAAATGGGTATGAG 290  
QY 125 GCATGTGCGTGAATGACGACCCCATGTCGACAGATGAAGACATTCATCAACAATAA 184  
Db 291 GCATGTGCGTGAATGACGACCCCATGTCGACAGATGAAGACATTCATCAACAATAA 350  
QY 185 AGGACATGTGACCCAGCATCTCTGATCTGTTGAAG 223  
Db 351 AGGAAATGTAAGTGAAGCTTCTACTTACCTGTTTCATG 389

RESULT 7  
AF001264 1265 bp mRNA MAM 19-NOV-1997  
LOCUS Bos taurus clone 4 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.  
ACCESSION AF001264  
NID 92623768  
VERSION AF001264.1 GI:2623768  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 1265)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
JOURNAL endothelial chloride channel  
AUTHORS Unpublished  
TITLE 2 (bases 1 to 1265)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA

FEATURES  
source Location/Qualifiers  
1. .1265  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/tissue\_type="lung"  
/cell\_type="aortic endothelial cells"  
109. .1137  
/note="endothelial adhesion molecule; chloride channel;  
secreted form"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86532.1"  
/db\_xref="PID:92623769"  
/db\_xref="GI:2623769"  
/translation="MVICLVNLIFFLHLPGKSSMNLINNGYDGIYAIINPSVE  
DKLEIKEMYTEASTLYFHATKRVRVFNYSILIPMTWKSSTYFTPKDSYQAD  
VIVANPYLYGDDPTLYGRCGEKRTIHTPNLITNNPHYSGRGVPTHEVAHL  
RWGIDETVNDQPFYISRNKTIETATCSHTIGINVFYKPCPGSCITSLCRDSOTG  
LYEACTFLPKRSQAKESIMPELSHSTVECTEKTHTNTEAPNLONKSNKSTWDY  
INNSVDFOPTSPTEMNPTHTPELSKQRVCLVDKSGMSAEDLIYLLALIKI  
FKLIGNTI"

BASE COUNT 440 a 231 c 231 g 363 t  
ORIGIN

Query Match 40.8%; Score 91; DB 3; Length 1265;  
Best Local Similarity 66.7%; Pred. No. 2.8e-17;

Matches 146: Conservative 0; Mismatches 70; Indels 3; Gaps 1;

QY 5 TCACAGGAGATGACAGCATGGGGCCATTAGAGTTCTGTCTGATCTTC 64  
 DB 89 TTACTGTAACTGTCAAAAGATGCTCTGTCTGAATGTTATTTCTTACTTTC 148  
 QY 65 ACCCTCTAGAGGGGCCCTGATTAATTCATTCAGTCAACAACTGGCTATGAG 124  
 DB 149 ATCTCTGCTGGG---AATGAAAGATTCATATGTAATTTGATTAACATGGGTATG 205  
 QY 125 GCATTGTGCTGCAATCGACCCCAATGTGCCAGAGATGAACACTCATTCACAAATA 184  
 DB 206 GCATTGTGCTGCAATTCACCCAGGTGCCAGAGATGAACAACTCATTCGAAACATTA 265  
 QY 185 AGACATGCTGACCCAGCATCTCTGTATCTGTGAG 223  
 DB 266 AGGAATGCTACTGAACTCTACTTACTTCTTCTATG 304

RESULT 8  
 BTU36445 2984 bp mRNA MAM 09-FEB-1996  
 LOCUS Bos taurus calcium-activated chloride channel mRNA, complete cds.  
 DEFINITION U36445  
 ACCESSION U36445  
 NID 91184065  
 VERSION U36445.1 GI:1184065  
 KEYWORDS  
 SOURCE cow.  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 Bovinae; Bos.  
 1 (bases 1 to 2984)  
 Cuningham, S.A., Awayda, M.S., Bubiou, J.K., Ismailov, I.I.,  
 Ariste, M.P., Bertier, B.K., Benos, D.J., and Fuller, C.M.  
 Cloning of an epithelial chloride channel from bovine trachea  
 J. Biol. Chem. 270 (52), 31016-31026 (1995)  
 96125078  
 2 (bases 1 to 2984)  
 Benos, D.J.  
 Direct Submission  
 Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,  
 University of Alabama at Birmingham, BHSB 706, Birmingham, AL  
 35294, USA

FEATURES  
 source  
 Location/Qualifiers  
 1..2984  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /cell\_type="epithelial cells"  
 /tissue\_type="trachea"  
 19..2730  
 /note="epithelial chloride channel"  
 /codon\_start=1  
 /product="calcium-activated chloride channel"  
 /protein\_id="AAC4851.1"  
 /db\_xref="PIR:G1184066"  
 /translation="MVRRLVILFLTHLPKSSMVLNNGIDGIVIAINPSVE  
 DEKLIQNKEMVTEASTYLEFATKRRVFRNVSILIPMTKSKSEYILMKOSYDAE  
 VIVANPYLKHGDDPTLYQGRGEGKGYIHETPNFLTNLPIYSGRGAFAVEMAH  
 RMGIFPEYNDGDPFYSRNTIEATRCSTHGTGNTIVYCOGSGCITRCSRDSOGL  
 YENKCTFPEKSTAKESIMFOSLSVTEFCEKHNTEAPLQMKMGKSTWDY  
 FLIOIIEKSLVGMVTFDSVAELRNMLKITDNYENTIANLPOEANGSTICRGLK  
 AGFOALISQOSTGSEIILLDGEDNEINSCIEYKOSGVIIHTIALGSPSAKELET  
 LDMTGHRFRANKDINGLIMAFSRISRSRSGSITOOTIOLESKALATKSPSAKELET  
 VDTSTGNDTFYVWTWTKIKKPEILLDPKGRKYSIDFKEDKLNHSAIRLIGIAETG  
 TWYSLNNHASPQILLYVTYTAASPTTPPYATATMSQNRHAHPSPIVIAVOSG  
 FLPLVGINVTAILETEDEGQVTLMDNGAGDTPANDGTSYTFIDYNGNRSYK  
 HAERNNVATLSLROPQNKALYIPGYIENKTIILNPREVKDILKALTEDESRITS  
 GSGFTVSGAPGNHPSVLPNPKIIDEAKKEDHIQLSWTAPANVLDRKANSTIIRI  
 SKSEFLDKQEDFNATLVNTSLKPKKSGSDNEFEPEFRLENGTNPFIYIAVOAINEA  
 NLSEVSNIAQATKIFPMEDSPALGKTISAINTALIFALMILTSIV"

CDS  
 1..2984  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /cell\_type="epithelial cells"  
 /tissue\_type="trachea"  
 19..2730  
 /note="epithelial chloride channel"  
 /codon\_start=1  
 /product="calcium-activated chloride channel"  
 /protein\_id="AAC4851.1"  
 /db\_xref="PIR:G1184066"  
 /translation="MVRRLVILFLTHLPKSSMVLNNGIDGIVIAINPSVE  
 DEKLIQNKEMVTEASTYLEFATKRRVFRNVSILIPMTKSKSEYILMKOSYDAE  
 VIVANPYLKHGDDPTLYQGRGEGKGYIHETPNFLTNLPIYSGRGAFAVEMAH  
 RMGIFPEYNDGDPFYSRNTIEATRCSTHGTGNTIVYCOGSGCITRCSRDSOGL  
 YENKCTFPEKSTAKESIMFOSLSVTEFCEKHNTEAPLQMKMGKSTWDY  
 FLIOIIEKSLVGMVTFDSVAELRNMLKITDNYENTIANLPOEANGSTICRGLK  
 AGFOALISQOSTGSEIILLDGEDNEINSCIEYKOSGVIIHTIALGSPSAKELET  
 LDMTGHRFRANKDINGLIMAFSRISRSRSGSITOOTIOLESKALATKSPSAKELET  
 VDTSTGNDTFYVWTWTKIKKPEILLDPKGRKYSIDFKEDKLNHSAIRLIGIAETG  
 TWYSLNNHASPQILLYVTYTAASPTTPPYATATMSQNRHAHPSPIVIAVOSG  
 FLPLVGINVTAILETEDEGQVTLMDNGAGDTPANDGTSYTFIDYNGNRSYK  
 HAERNNVATLSLROPQNKALYIPGYIENKTIILNPREVKDILKALTEDESRITS  
 GSGFTVSGAPGNHPSVLPNPKIIDEAKKEDHIQLSWTAPANVLDRKANSTIIRI  
 SKSEFLDKQEDFNATLVNTSLKPKKSGSDNEFEPEFRLENGTNPFIYIAVOAINEA  
 NLSEVSNIAQATKIFPMEDSPALGKTISAINTALIFALMILTSIV"

FEATURES  
 source  
 Location/Qualifiers  
 1..3288  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /cell\_type="aortic endothelial cells"  
 /clone="2"  
 /tissue\_type="lung"  
 59..2443  
 /note="endothelial adhesion molecule; chloride channel"  
 /codon\_start=1  
 /product="Lu-ECAM-1"  
 /protein\_id="AAB86530.1"  
 /db\_xref="PIR:G2623765"  
 /db\_xref="GI:2623765"

BASE COUNT 1027 a 583 c 561 g 813 t  
 ORIGIN

Query Match 40.1%; Score 89.4; DB 3; Length 2984;  
 Best Local Similarity 67.3%; Pred. No. 9.1e-17;  
 Matches 142: Conservative 0; Mismatches 66; Indels 3; Gaps 1;

QY 13 AGATGTACAGCAATGGGGCCATTAAAGATTCTGTGATCTTGAATCTTCTTA 72  
 DB 7 ACATGGCCAAATAATGGGCTCGTCGATGCTGATCTGTCTTACTTGGCACCTCTG 66  
 QY 73 GAAGGGCCCTGAGTAATTTACTCATTTACCTTAACAACAATGGCATGAAGCATGTC 132  
 DB 67 CCTGG---AATGAAAGATTAAATGTAATTTATTAACATGATGATGAGCATTTGTC 123  
 QY 133 GTTGCAATCGACCCCAATGTGCCAGAGATGAACACTCATTCACAAATAAAGACATG 192  
 DB 124 ATTCAATTAACCTTACTGATGTGCCAGAGATGAACAACTCATTCACAAATAAAGAAATG 183  
 QY 193 GTGACCCAGCATCTCTGTATCTGTTGAAG 223  
 DB 184 GTACTGAGCTTCTACTTACTTCTTCTTCTATG 214

RESULT 9  
 AF001262 3288 bp mRNA MAM 19-NOV-1997  
 LOCUS Bos taurus clone 2 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF001262  
 NID 92623764  
 VERSION AF001262.1 GI:2623764  
 KEYWORDS  
 SOURCE Bos taurus.  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 Bovinae; Bos.  
 1 (bases 1 to 3288)  
 Eblie, R.C., Widom, J., Gruber, A.D., Abdel-chany, M., Levine, R.,  
 Goodwin, A., and Pauli, B.U.  
 Cloning and characterization of Lu-ECAM-1 suggest it is an  
 endothelial chloride channel  
 unpublished  
 2 (bases 1 to 3288)  
 Eblie, R.C., Widom, J., Gruber, A.D., Abdel-chany, M., Levine, R.,  
 Goodwin, A., and Pauli, B.U.  
 Direct Submission  
 Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
 Ithaca, NY 14853, USA

FEATURES  
 source  
 Location/Qualifiers  
 1..3288  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /cell\_type="aortic endothelial cells"  
 /clone="2"  
 /tissue\_type="lung"  
 59..2443  
 /note="endothelial adhesion molecule; chloride channel"  
 /codon\_start=1  
 /product="Lu-ECAM-1"  
 /protein\_id="AAB86530.1"  
 /db\_xref="PIR:G2623765"  
 /db\_xref="GI:2623765"

CDS  
 1..3288  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /cell\_type="aortic endothelial cells"  
 /clone="2"  
 /tissue\_type="lung"  
 59..2443  
 /note="endothelial adhesion molecule; chloride channel"  
 /codon\_start=1  
 /product="Lu-ECAM-1"  
 /protein\_id="AAB86530.1"  
 /db\_xref="PIR:G2623765"  
 /db\_xref="GI:2623765"

[illegible]

BASE COUNT	923 a	710 c	661 g	728 t
ORIGIN	HPDLQEDFNNTLYNASSLLPKKAGSGSETEKFKPRTFKINGIQLYIAIADNEASL TSEVNIQAQAVKLSLEDISIALGDSDAISMTIMGLTVINSILN"			

Query Match	32.48;	Score 72.2;	DB 12;	Length 3022;
Best Local Similarity	71.48;	Pred. No. 1,2e-11;		
Matches 95;	Conservative 0;	Mismatches 38;	Indels 0;	Gaps 0;

Qy	91	TCACGATTCAGCTCAACAACAAATGGCATTAAGCATATGCTGTGAATCGACCCCAAT	150
Db	81	TCGATGTCATCTCAACAGCAATGATGATACGAGGGTGTGTGCATTCGCAATTAACCCAGT	140
Qy	151	GTGCCGGAAGATGAAACACCTCATTCACAAATAATAAGACATGTGACCCAGGATCTCTG	210
Db	141	GTGCCGAGAGACGAAAGAGCTCATTCGACAGCATTAAGAAATGTGTAAGCTCAAGCCTTACC	200
Qy	211	TATCTGTTGAAG	223
Db	201	TACTGTGTTGAAG	213

RESULT 12  
AF052746

LOCUS	3471 bp	MRNA	ROD	08-APR-1999
DEFINITION	Mus musculus chloride channel cacc mRNA, complete cds.			

ACCESSION AF052746  
NID 93560546  
VERSION AF052746.1 GI:3560546

**KEYWORDS** house mouse.

ORGANISM      Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciuriformes; Muridae; Mus

REFERENCE AUTHORS	1 (bases 1 to 3471)  Romfo, J., Musante, P., Cinti, R., Serini, M., Moran, O., Zedaira-Moran, O., Echeverria, A., Rodenas, I., Scruognaletti, M., Linares, M., Linares, M.
----------------------	--

**TITLE** and Galietta, L.J.V.  
Characterization of a murine gene homologous to the bovine CAC

chloride channel  
Gene 228 (1-2), 181-188 (1999)  
JOURNAL  
MEDICINE 98173882

REFERENCE AUTHORS	PUBLICATION YEAR
2 (bases 1 to 3471)	1961/1962
Romio, L., Musante, L., Cinti, R., Moran, O., Serl, M. and	

Galietta, L.J.V.  
Direct Submission  
TITLE  
FOOTNOTES  
REFERENCES

**FEATURES**  
JOURNAL  
Submitted (09-MAR-1998) Lab. di Genetica  
Giannina Gaslini, Genova I-16148, Italy  
Location/Qualifiers

Source

CDS

```

source
1. 3471
/organism="Mus musculus"
/db_xref="taxon:10090"
/db_xref="dbEST:AA245677"
47. 2752
/codon_start=1
/product="chloride channel Cacc"
/protein_id="AAC35003.1"
/db_xref="PDB:93560547"
/db_xref="GI:3560547"
/translaton="MVEGLQVLEFLTILHLQNTESMVHLNSNGEVLVAINPSVEP
DERILIPSTIKEMVQASTYLEFASOGRYVFNISLIVTWTWSKPEYLMPKESDCKAD
VIVADPHLQHGDDPYLTLOYGCGGRGVQIHTPTNLLDNLRIYGPGRVHVMAHL
RMGEVDENNDQPPYMTKRTITBATRGSTRITGVNVCNRCGVQIRACGRSDSTL
YEPCEFTIPDKIQTAGASINFMQNLNVEECTCKNNHAEVNLQNKCNKRSTWDYI
KTSDFQNAFPMRGTEAPPEPTFSLSRRKVALVDKSSMKRDERLIMNNAEL
YLTVIVERKSNVGLTPEPSAHLQNYLILKITSDDYOKITNLPQAAGGTSICGHQ
AGFAITSSDDSGSEIVLTDDEGNDIRSCFEAVSRGAIHTITAGSAAELETFE
LSDMTGSGRFYAKNDLNSLIDAFSRISGTSVGOALOLESKAFVRAQAWINGTVP
LDSYVNGDTPEVITWYWKPEIILIDPGRKKTYSDDODKLTNSARLQIPGAELG
TMYSTIGKTSQLTITMTVTARSPMEPLLAATHMSQSTQYPSRMIYARVSGQL
PVLANTATLIEAHGHQVTLLEMDNGAGATVANDGTYITFTDIGHNSKISLKV
QAQNCNRRLSLRQKNSLYIPGVENGLINPNRPVQVEALATVEDENRYTSGSS
FTVSGAPGDGHARVPPSKVTDLAETIGYIHLTAPGVLVNDGNRAHYIRIMSO

```

BASE COUNT	1096 a	777 c	736 g	862 t
ORIGIN	HPDLDQDPNNATLVNASSLIPKEAGSKKEFFKRPPEFFKLANGQLYIAQADNEASL TSEVSNIAQAVKLTLSLEDSISALGDISAICMTIWLGLVINSILN"			

Query Match	32.4%;	Score 72.2;	DB 12;	Length 3471;
Best Local Similarity	71.4%;	Pred. No. 1.2e-11;		
Matches	95;	Conservative	0;	Mismatches 38;
			Indels	0;
			Gaps	0;

QY	91	TCACCATTCAGCTAAACAACAATGGCTATTAAGCAATGTGTCATGTAGCCCCAAT	150
Db	110	TCCATGTGTGATCTCTCAACAGCAATGATAGCAGGGTGTGTGTCATTGCCATTAAACCCAGT	160
QY	151	GTGCCAGAAATATGAACACCTCATTCACAAATTAAGACATGTGTACCCAGGCACTCTG	210
Db	170	GTCCGGAGAGCAAAAGGCTCATCCACAGCATTAAGGAAATGTACTCAAGGCTTACC	220
QY	211	TATCTGTGTGAAG	223
Db	230	TACCTGTTTGAAG	242

RESULT 13  
AB026833

LOCUS	AB026833	3604 bp	MRNA	PRI	26-MAY-1999
DEFINITION	Homo sapiens mRNA for chloride channel protein, complete cds.				

ACCESSION	AB026833
NID	94887600
VERSION	AB026833.1
	GT:4887600

**KEYWORDS** chloride channel protein.  
**SOURCE** Homo sapiens adult corneal epithelium cDNA to mRNA

ORGANISM	clone_lib:lamba zap clone:lamba 13802.
Homo sapiens	
Tubulastrea nutans	
Chordata	
Cnidaria	
Mollusca	
Vertebrata	
Mammalia	

REFERENCE  
1 (sites)  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS	TITLE
Itoh, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.	Isolation and characterization of chloride channel in human corneal

epithelium  
JOURNAL  
Unpublished (1999)  
REFERENCE  
2 (pages 1 to 3604)

REFERENCE  
2 (Pages 1 to 3004)  
AUTHORS  
Itoh, R., Kawamoto, S. and Okubo, K.  
TITLE  
Direct Submission

Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases. Reiko Itoh, Institute for Molecular and Cellular Biology, Osaka University, 1-1 Yamadaoka, Suita, Osaka 565-0871, Japan.

University; Yamada-oka 1-3, Suita, Osaka 565-0871, Japan  
(E-mail: reiko@imcb.osaka-u.ac.jp, Tel: 81-6-6879-7992  
Fax: 81-6-6877-1922)

**FEATURES**

Location/Qualifiers

FEATURES	Source	Location/Qualifiers
		1..3604
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="1"
		/clone="lambda 13802"
		/clone_1bp="lambda zap"
		/dev_stage="adult"
		/map="1p32"
		/tissue_type="corneal epithelium"
	CDS	42..2873
		/codon_start=1
		/product="chloride channel protein"
		/protein_id="BA07810.1"
		/db_xref="PID:d1041578"
		/db_xref="PID:g4887601"
		/db_xref="GI:4887601"
		/translation="MTORSIAGPCICNLKEFVTLILVALSEELPFLGAGVQLDONGYNGILL IAINPQVENONLISNKEMITEAFSEFLNATKRVFEKRIILIPATMANNSKIK OYEHKRNALVTDMYCGAGDDPYTLIOGCGKEGKYIHFPNFIANDNLGAGSRGR FSEYERAHNLFWGVDFEYDINDKPFYINGQNOIKYTRRSSDITGIFYCEKGCPOENDCII SKLFEBCETIYNSTONATASIMFQSLSSVFECSASTHNDAPNIAQNMOCGRSKW AETLMDADFHSFSPMGSTELPPPTFSLVAGDGVVCLVDVSSKMAEADRLQLQQA AEVTLQIVIHTEFVGIASDSKGEIRAOHQINSDNDRLIVSYSLPTVSATDTDISI





gene 1792.. 2217  
/gene="A3R"  
1792.. 2217  
/gene="A3R"  
/note="contains aminoacyl-tRNA synthetase class-II signature; PBCV-1 terminal repeat ORF, corresponds to GenBank Accession Number M55319"  
/protein\_id="AAC96371.1"  
/codon\_start=1  
/db\_xref="GI:624054"  
/db\_xref="GI:624054"  
/translation="MRDQTLYIKMSGNDPIFRFLKSFIFLLISLYNGLSNREGVDHKGGAAYDLCMFGEELLNRLNRTSLIGASWGVDRIGLNHDGLVGEVDELDGNHCEGSEGRDTHLQNRNRIYTGRRHFHSSFFPEYLNRHFYK"  
complement(1891.. 2127)  
/gene="A4L"  
complement(1891.. 2127)  
/gene="A4L"  
/note="PBCV-1 terminal repeat ORF; corresponds to GenBank Accession Number M55319"  
/codon\_start=1  
/protein\_id="AAC96372.1"  
/db\_xref="PID:9624091"  
/db\_xref="GI:624091"  
/translation="MSVSFTRAFTRFISAKNLFTNDRIYETTEPIDPPIPKTSSIQSIKFFETHTQLRVGTGFVNTPEIPEVDK"  
2288.. 3076  
/gene="A5R"  
2288.. 3076  
/gene="A5R"  
/note="contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157"  
/codon\_start=1  
/protein\_id="AAC96373.1"  
/db\_xref="PID:9624055"  
/db\_xref="GI:624055"  
/translation="KVNKCSKFTFAENGHDVCLKTLIEAGAPFDNGSGMTEHYAHTHPTACRKLIDAGANIDITDNGSCPIHRAVYNGDACYKLIVGAGALDVIDDCTGMPHHAAVYIGYDACVKMLIEAGALNIDGGIAPLRYAYKGHDVCALLYAGAPRLDITISGCTPLHRAVPEVGHDCASMLVNAKIVSEPLSCVPIQTSALVGLVLTITQIGRSEPAKITIAHLPVGARDLRTMLCLNRPWSRETSIIA"  
complement(2393.. 2611)  
/gene="A6L"  
complement(2393.. 2611)  
/gene="A6L"  
complement(2393.. 2611)  
/gene="A6L"  
/codon\_start=1  
/protein\_id="AAC96374.1"  
/db\_xref="PID:9624092"  
/db\_xref="GI:624092"  
/translation="MOMHGSSIINDVKSOTCDEQLHICMAIKNRMTKKWCASIIIDDVNICRINHELYARSKITINRMYOGPP"  
complement(3682.. 4242)  
/gene="A7L"  
complement(3682.. 4242)  
/gene="A7L"  
complement(3682.. 4242)  
/note="contains 4 ankyrin repeats; similar to D. melanogaster notch protein, corresponds to Swiss-Prot Accession Number P07027"  
/codon\_start=1  
/protein\_id="AAC96375.1"  
/db\_xref="PID:9624056"  
/db\_xref="GI:624056"  
/translation="MOEONLILLLELAGETPLHLAIRGLETCVGLINAGANLVNRNBRPTLHLAVICGGDANCILLIKAGALATNVRIDGAPTPTIASRLGHTIYVKEFLACGDEPLHFRKNDSTPLHFRNGKTKICAEELLITAGVEINTINNESATPLHFRASRARRRVVRDRIAGTLPISTIVPKHHYIG"  
complement(4285.. 4698)  
/gene="A8L"  
complement(4285.. 4698)  
/gene="A8L"  
complement(4285.. 4698)  
/note="contains 3 ankyrin repeats; similar to human erythrocyte ankyrin, corresponds to Swiss-Prot Accession

Query Match	Best Local Similarity	Score	DB 17:	Length	330742:
Matches 82; Conservative	54.3%;	Pred. No. 0.046;	Mismatches 69;	Indels 0;	Gaps 0;
gene					
CDS					
Number P16157"	/codon_start=1				
	/protein_id="AAC96376.1"				
	/db_xref="PID:g624057"				
	/db_xref="GI:624057"				
	/translation="MNLRKSNQERYKWTIRNLVTIGENINIGKSYENPPLTAVRRG				
	YTEVKELINAGANLAVRDEGRPLHAIIEWTICFILLNAGANLNGKKGFAP				
	LHAIATKCTDYMVLLINAGANLNDVGYSKGRHTY"				
	5208..5735				
	/gene="A9R"				
	5208..5735				
	/gene="A9R"				
	/codon_start=1				
	/protein_id="AAC96377.1"				
Query Match	18.2%;	Score 40.6;	DB 17:	Length 330742:	
Best Local Similarity	54.3%;	Pred. No. 0.046;	Mismatches 69;	Indels 0;	Gaps 0;
Matches 82; Conservative	0;	Mismatches 69;	Indels 0;	Gaps 0;	
gene					
CDS					
Number P16157"	/codon_start=1				
	/protein_id="AAC96376.1"				
	/db_xref="PID:g624057"				
	/db_xref="GI:624057"				
	/translation="MNLRKSNQERYKWTIRNLVTIGENINIGKSYENPPLTAVRRG				
	YTEVKELINAGANLAVRDEGRPLHAIIEWTICFILLNAGANLNGKKGFAP				
	LHAIATKCTDYMVLLINAGANLNDVGYSKGRHTY"				
	5208..5735				
	/gene="A9R"				
	5208..5735				
	/gene="A9R"				
	/codon_start=1				
	/protein_id="AAC96377.1"				

## COMMENT

Baylor Plaza, Houston, TX 77030, USA  
On Jan 31, 1999 this sequence version replaced gi:4204112.  
INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email  
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

## ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.

Version: 1.01 gxfco.

## Location/Qualifiers

FEATURES  
source

1. .135039  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RPC14-736H12"  
/chromosome="3p22-8"  
/complement(1. .1909)  
/rpt\_family="LIMA"  
repeat\_region  
complement(1916. .2345)  
/rpt\_family="MER65\_internal"  
repeat\_region  
complement(2490. .2802)  
/rpt\_family="MER31\_internal"  
repeat\_region  
complement(2810. .2928)  
/rpt\_family="AluSg/x"  
repeat\_region  
complement(2937. .3164)  
/rpt\_family="Aluub"  
repeat\_region  
complement(3167. .3367)  
/rpt\_family="MER65\_internal"  
repeat\_region  
3400. .3690  
/rpt\_family="AluSx"  
complement(3754. .4032)  
/rpt\_family="MER65\_internal"  
repeat\_region  
complement(3930. .4220)  
/rpt\_family="MER31\_internal"  
repeat\_region  
4222. .4522  
/rpt\_family="Aluub"  
repeat\_region  
4543. .4582  
/rpt\_family="AT-rich"  
complement(4598. .4896)  
/rpt\_family="MER4\_internal"  
repeat\_region  
complement(4973. .5286)  
/rpt\_family="MER4\_internal"  
repeat\_region  
complement(5359. .5841)  
/rpt\_family="AluY"  
repeat\_region  
complement(5843. .6300)  
/rpt\_family="MER4\_internal"  
repeat\_region  
complement(6354. .6383)

repeat\_region  
complement(6452. .6479)  
/rpt\_family="AT-rich"  
repeat\_region  
complement(6510. .6726)  
/rpt\_family="LIPB3"  
repeat\_region  
complement(6727. .6997)  
/rpt\_family="AluSx"  
repeat\_region  
complement(7010. .7805)  
/rpt\_family="LIMA3"  
repeat\_region  
7806. .8068  
/rpt\_family="AluY"  
complement(8302. .9621)  
/rpt\_family="MER65\_internal"  
repeat\_region  
complement(9617. .9703)  
/rpt\_family="MER65\_internal"  
repeat\_region  
complement(9688. .9818)  
/rpt\_family="MER65\_internal"  
repeat\_region  
complement(9812. .10133)  
/rpt\_family="MER65\_internal"  
10195. .10319  
/rpt\_family="Aluub"  
repeat\_region  
complement(10360. .10764)  
/rpt\_family="MER65\_internal"  
repeat\_region  
complement(11044. .11485)  
/rpt\_family="MER65A"  
11486. .11828  
/rpt\_family="MTIC"  
13746. .14256  
/note="region: Similar to q918c09.x1 Homo sapiens CDNA,  
A1334028."  
14327. .14677  
/note="region: Similar to qe71d09.x1 Homo sapiens CDNA,  
A1193625."  
complement(14678. .14978)  
/rpt\_family="AluSx"  
14978. .15110  
/note="region: Similar to qm24h01.x1 Homo sapiens CDNA,  
A1279296."  
15079. .15279  
/note="region: Similar to aJ52h07.s1 Homo sapiens CDNA,  
AA854044."  
15348. .15571  
/note="region: Similar to nk51g07.s1 Homo sapiens CDNA,  
AA546981."  
16847. .17312  
/rpt\_family="L2"  
complement(18385. .18675)  
/rpt\_family="AluSg"  
repeat\_region  
complement(18840. .18867)  
/rpt\_family="AT-rich"  
repeat\_region  
complement(18874. .19069)  
/rpt\_family="LIME3A"  
19557. .19811  
/rpt\_family="MT1A1"  
complement(19812. .20250)  
/rpt\_family="MT1C"  
20355. .20645  
/rpt\_family="MT1C"  
20679. .20714  
/rpt\_family="U2"  
21188. .22073  
/rpt\_family="LIM4"  
22103. .22244  
/rpt\_family="AluSg/x"  
22246. .22402  
/rpt\_family="LIPB8"  
22433. .22613  
/rpt\_family="AluY"  
22619. .22921  
/rpt\_family="AluSg"  
complement(23264. .23330)  
/rpt\_family="MER5A"  
23831. .23871  
repeat\_region



Gencore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 07:38:49 ; Search time 3059.95 Seconds  
(without alignments)  
143.752 Million cell updates/sec

Title: US-09-049-696-1  
Perfect score: 223  
Sequence: 1 GAAATCACAGGAGATGTAC.....ATCTCTATCTCTTGAAG 223

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 98626752 residues

Database : EST.\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	26	11.7	428	30	284017	284017 SS284017 Po
2	19	8.5	688	42	A1133878	A1133878 GH10661.5
3	19	8.5	593	46	A1455107	A1455107 LD21194.3
4	17	7.6	267	20	D20237	D20237 HUNGSO1211
5	17	7.6	407	21	T62598	T62598 yC04b11.r1
6	17	7.6	264	22	R28408	R28408 yH50b07.r1
7	17	7.6	82	22	R69402	R69402 yJ83d10.r1
8	17	7.6	489	22	R75771	R75771 y121a10.r1
9	17	7.6	291	23	F14976	F14976 SSC7H01 Por
10	17	7.6	352	23	H31005	H31005 SM31CA272SK
11	17	7.6	492	23	H43477	H43477 y064e06.r1
12	17	7.6	344	24	D18308	D18308 M05GS01184
13	17	7.6	581	24	H74066	H74066 yS13g08.r1
14	17	7.6	580	24	N21299	N21299 yX53e08.s1
15	17	7.6	537	24	N28457	N28457 yX35g05.r1
16	17	7.6	341	25	N72239	N72239 zA01f07.s1
17	17	7.6	493	25	N80448	N80448 zA15b12.r1
18	17	7.6	469	25	N95656	N95656 zA29e08.r1
19	17	7.6	282	26	W52109	W52109 zC48c06.r1
20	17	7.6	533	26	W97675	W97675 mF98b01.r1
21	17	7.6	421	27	AA001230	AA001230 zH82g03.r
22	17	7.6	535	27	AA053411	AA053411 z171c07.s
23	17	7.6	629	29	W99845	W99845 mG28h11.r1
24	17	7.6	629	29	AA142696	AA142696 mR8e10.r
25	17	7.6	744	30	AA198886	AA198886 mU16g02.r
26	17	7.6	475	31	AA284073	AA284073 zS49a11.s
27	17	7.6	409	33	AA412477	AA412477 zT95g10.s
28	17	7.6	406	33	AA445980	AA445980 zW59e11.r
29	17	7.6	500	34	AA75187	AA75187 vG97h05.r
30	17	7.6	450	34	AA510978	AA510978 vG38a09.r
31	17	7.6	320	37	AA668949	AA668949 aB92b11.s
32	17	7.6	513	37	AA707398	AA707398 zJ27g02.s
33	17	7.6	379	38	AA775297	AA775297 aC77g04.s
34	17	7.6	454	39	AA887216	AA887216 cJ53c03.s
35	17	7.6	413	40	AA978118	AA978118 cP32a04.s
36	17	7.6	414	41	A1019560	A1019560 uA90d02.r
37	17	7.6	376	41	A1019833	A1019833 uA91g08.r
38	17	7.6	478	43	A1160519	A1160519 qC82f12.x
39	17	7.6	724	45	A1324543	A1324543 mR8e10.y
40	17	7.6	327	45	A1356791	A1356791 gY17g03.x
41	17	7.6	395	45	A1383455	A1383455 tE30f08.x
42	17	7.6	431	47	A1477984	A1477984 tD49h10.x
43	17	7.6	693	48	A1559441	A1559441 tQ33h02.x
44	17	7.6	584	48	A1572665	A1572665 tG68d03.x
45	17	7.6	442	51	A1736995	A1736995 sB36d03.y

# ALIGNMENTS

RESULT	1
284017	284017
LOCUS	428 bp mRNA
DEFINITION	SS284017 Porcine small intestine cDNA library Sus scrofa cDNA clone
ACCESSION	C12a03 5', mRNA sequence.
NID	284017
VERSION	284017.1 GI:1806336

```

KEYWORDS      EST.
SOURCE         pig.
ORGANISM       Sus scrofa
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE      1 (bases 1 to 428)
AUTHORS        Wintero,A.K., Fredholm,M. and Davies,W.
TITLE          Evaluation and characterization of a porcine small intestine cDNA
               library: analysis of 839 clones
JOURNAL        Mamm. Genome 7 (7), 509-517 (1996)
MEDLINE        9637607
COMMENT        On Sep 12, 1996 this sequence version replaced gi:1404772.

FEATURES
  source       1. 428
               /organism="Sus scrofa"
               /db_xref="taxon:9623"
               /clone="C12a03"
               /clone_11b="Porcine small intestine cDNA library"
               /note="directionally cloned cDNA in X11-blue MRF"

BASE COUNT    134 a      87 c      100 g      102 t      5 others

ORIGIN
Query Match   11.7%; Score 26; DB 30; Length 428;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GTGCCAGAGATGAACACTCATTC A 176
|||||
DB 152 GTGCCAGAGATGAACACTCATTC A 177

RESULT 2
A1133878      688 bp      mRNA      EST      18-NOV-1998
LOCUS        GH10661.5prime GH Drosophila melanogaster head pot2 Drosophila
DEFINITION   melanogaster cDNA clone GH10661 5prime, mRNA sequence.
ACCESSION    A1133878
VERSION      93626436
KEYWORDS     EST.
SOURCE       fruit fly.
ORGANISM     Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
               Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
               Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE     1 (bases 1 to 688)
AUTHORS      Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
               Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE        BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (1997)
COMMENT      Contact: Harvey, D.
               G. M. Rubin-Molecular and Cell Biology
               University of California Berkeley
               539 USA, Berkeley, CA 94720-3200, USA
               Fax: 510 643 9947.
               Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
               Plate: 106 row: F column: 1
               High quality sequence stop: 529.
               Location/Qualifiers
               1..688
               /organism="Drosophila melanogaster"
               /db_xref="taxon:7227"
               /clone="GH10661"
               /clone_11b="GH Drosophila melanogaster head pot2"
               /sex="male and female"
               /dev_stage="adult"

```

```

/1ab_host="DH5 - alpha"
/note="Organ: head; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."

BASE COUNT    201 a      178 c      131 g      178 t

ORIGIN
Query Match   8.5%; Score 19; DB 42; Length 688;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 ATGAACACTCATTCAC A 179
|||||
DB 240 ATGAACACTCATTCAC A 258

RESULT 3
A1455107/c    593 bp      mRNA      EST      09-MAR-1999
LOCUS        LD21194.3prime LD Drosophila melanogaster embryo pot2 Drosophila
DEFINITION   melanogaster cDNA clone LD21194 3prime, mRNA sequence.
ACCESSION    A1455107
VERSION      94299645
KEYWORDS     EST.
SOURCE       fruit fly.
ORGANISM     Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
               Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
               Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE     1 (bases 1 to 593)
AUTHORS      Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
               Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE        BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (1997)
COMMENT      On May 7, 1998 this sequence version replaced gi:1121029.
               Other ESTs: LD21194.5prime
               Contact: Harvey, D.
               G. M. Rubin-Molecular and Cell Biology
               University of California Berkeley
               539 USA, Berkeley, CA 94720-3200, USA
               Fax: 510 643 9947
               Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
               Based upon the presence of a XhoI site followed by a run of 14 or
               more T residues at the beginning of the sequence, this clone was
               polyadenylated. The resulting Poly-T sequence has been removed.
               Plate: 211 row: H column: 10
               High quality sequence stop: 498.
               Location/Qualifiers
               1..593
               /organism="Drosophila melanogaster"
               /db_xref="taxon:7227"
               /clone="LD21194"
               /clone_11b="LD Drosophila melanogaster embryo pot2"
               /sex="male and female"
               /dev_stage="0 to 24 hours mixed stage embryonic"
               /1ab_host="X11 Blue"
               /note="Organ: embryo; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2."

BASE COUNT    152 a      104 c      145 g      192 t

ORIGIN
Query Match   8.5%; Score 19; DB 46; Length 593;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 ATGAACACTCATTCAC A 179
|||||
DB 453 ATGAACACTCATTCAC A 435

```

RESULT 4  
 D20237/c 267 bp mRNA EST 30-JUL-1996  
 LOCUS HOMGS01211 Human promyelocyte Homo sapiens cDNA clone pml277 3',  
 DEFINITION mRNA sequence.  
 ACCESSION D20237  
 NID 9501334  
 VERSION D20237.1 GI:501334  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 267)  
 Okubo, K., Fukushima, A., Yoshii, J., Miyama, T., Kojima, Y.,  
 Yoshinari, H., Arimoto, J. and Matsubara, K.  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 3-1 Yamada-oka, Suita, Osaka 565, Japan.  
 Location/Qualifiers  
 1. .267  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="pml277"  
 /note="Human promyelocyte"  
 /note="Female, adult, cell\_line = HL60, cell\_type =  
 promyelocyte."  
 BASE COUNT 82 a 33 c 47 g 94 t 11 others  
 ORIGIN  
 Query Match 7.6%; Score 17; DB 20; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ATTCTCACCCTCTAGA 74  
 ||||||||||||||||  
 Db 19 ATTCTCACCCTCTAGA 3

RESULT 5  
 T62598/c 407 bp mRNA EST 16-FEB-1995  
 LOCUS T62598  
 DEFINITION YC04911.r1 Strata gene lung (#937210) Homo sapiens cDNA clone  
 IMAGE:79677 5', mRNA sequence.  
 ACCESSION T62598  
 NID 9666255  
 VERSION T62598.1 GI:666255  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 407)  
 Hillier, L., Lennon, G., Becker, M., Donald, M.F., Chippelli, B.,  
 Chissole, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
 and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 JOURNAL Genome Res. 6 (9), 807-828 (1996)  
 COMMENT 97044478

CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 Insert size: 632  
 High quality sequence stops: 293 Source: IMAGE Consortium, LNL This  
 clone is available royalty-free through LNL; contact the IMAGE  
 Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 632 Std Error: 0.00  
 Seq primer: M13RP1  
 High quality sequence stop: 293.  
 Location/Qualifiers  
 1. .407  
 /organism="Homo sapiens"  
 /db\_xref="GDB:483294"  
 /db\_xref="taxon:9606"  
 /map="2"  
 /clone="IMAGE:79677"  
 /clone\_id="Stratagene lung (#937210)"  
 /sex="male"  
 /dev\_stage="72 years"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: lung; Vector: pBluescript SK-; Site: 1:  
 EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. normal lung. Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GATTTCGCGCAG  
 3' -3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGT 3'."

BASE COUNT 121 a 52 c 80 g 152 t 2 others  
 ORIGIN  
 Query Match 7.6%; Score 17; DB 21; Length 407;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ATTCTCACCCTCTAGA 74  
 ||||||||||||||||  
 Db 99 ATTCTCACCCTCTAGA 83

RESULT 6  
 R28408/c 264 bp mRNA EST 25-APR-1995  
 LOCUS R28408  
 DEFINITION YH50D07.r1 Soares placenta Nb2HP Homo sapiens cDNA clone  
 IMAGE:133141 5', mRNA sequence.  
 ACCESSION R28408  
 NID 9784543  
 VERSION R28408.1 GI:784543  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 264)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 The WASHU-MERCK EST Project  
 Unpublished (1995)  
 TITLE The WASHU-MERCK EST Project  
 JOURNAL  
 COMMENT  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 Insert size: 984  
 High quality sequence stops: 121  
 Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 984 Std Error: 0.00  
 Seq primer: M13Rp1  
 High quality sequence stop: 121.

## FEATURES

## source

Location/Qualifiers

1.264  
 /organism="Homo sapiens"  
 /db\_xref="GDB:538822"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:133141"  
 /clone\_1ib="Soares placenta Nb2hp"  
 /sex="Female"  
 /dev\_stage="Placenta obtained at birth (full term)"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGAGAGATTCGCGCGGAGAGATTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

## BASE COUNT

76 a 32 c 60 g 95 t 1 others

## ORIGIN

## Query Match

Best Local Similarity 7.6%; Score 17; DB 22; Length 264;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ATTCTTCACCTCTAGA 74  
 |||  
 DB 36 ATCTTCACCTCTAGA 20

## RESULT 7

## R69402/c

## LOCUS

## DEFINITION

## KEYWORDS

## ACCESSION

## NID

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

R69402 82 bp mRNA EST 01-JUN-1995  
 YJ83d10.r1 Soares breast 2NBHst Homo sapiens cDNA clone  
 IMAGE:155347 5', similar to gb:U74715 ALPHA-L-1DROXIDASE PRECURSOR  
 (HUMAN);, mRNA sequence.  
 R69402  
 9842919  
 R69402.1 GI:842919  
 EST.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 82)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995).  
 On May 8, 1995 this sequence version replaced gi:800646.  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 Insert Size: 750  
 High quality sequence starts: 1  
 High quality sequence stops: 1  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.

Trace considered overall poor quality  
 Insert Length: 750 Std Error: 0.00  
 Seq primer: M13Rp1  
 High quality sequence stop: 1.

## FEATURES

## source

Location/Qualifiers

1.82  
 /organism="Homo sapiens"  
 /db\_xref="GDB:572254"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:155347"  
 /clone\_1ib="Soares breast 2NBHst"  
 /sex="Female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACACATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot - 230. Library constructed by Bento Soares and M.Fatima Bonaldo."

## BASE COUNT

16 a 39 c 13 g 13 t 1 others

## ORIGIN

## Query Match

Best Local Similarity 7.6%; Score 17; DB 22; Length 82;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GAAGCATTCGCTGC 137  
 |||  
 DB 51 GAAGCATTCGCTGC 35

## RESULT 8

## R75771

## LOCUS

## DEFINITION

## KEYWORDS

## ACCESSION

## NID

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

R75771 489 bp mRNA EST 06-JUN-1995  
 YJ21a10.r1 Soares breast 2NBHst Homo sapiens cDNA clone  
 IMAGE:158874 5', mRNA sequence.  
 R75771  
 9850453  
 R75771.1 GI:850453  
 EST.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 489)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 Insert Size: 713  
 High quality sequence stops: 376  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 713 Std Error: 0.00  
 Seq primer: M13Rp1  
 High quality sequence stop: 378.



FEATURES  
source  
1. .489  
/organism="Homo sapiens"  
/db\_xref="GDB:572927"  
/db\_xref="taxon:9606"  
/clone="IMAGE:158874"  
/clone.lib="Soares breast 2NBH8t"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTACCAATGAGAGGAGCGCCCTTTTCTTTTCTTTT 3'), TGTACCAATGAGAGGAGCGCCCTTTTCTTTTCTTTT 3'), double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 230. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 119 a 80 c 140 g 140 t 10 others  
ORIGIN

Query Match 7.6%; Score 17; DB 22; Length 489;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 CACTCATCAACAATA 183  
|||||  
Db 272 CACTCATCAACAATA 288

RESULT 9  
F14976 291 bp mRNA EST 09-SEP-1996  
LOCUS SSC7001 Porcine small intestine cDNA library Sus scrofa cDNA clone  
DEFINITION c7h01, mRNA sequence.  
ACCESSION F14976  
NID 9972201  
VERSION F14976.1 GI:972201  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
REFERENCE  
1 (bases 1 to 291)  
Wintero, A.K., Fredholm, M. and Davies, W.  
Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones  
Mamm. Genome 7 (7), 509-517 (1996)  
JOURNAL  
MEDLINE  
COMMENT  
On May 8, 1995 this sequence version replaced gi:799603.  
Contact: A.K. Wintero  
Department of Animal Science and Animal Health, Division of Animal Genetics, The Royal Veterinary and Agricultural University  
Bulowsvej 13, 1870 Frederiksberg C, Denmark.

FEATURES  
source  
1. .291  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone="c7h01"  
/clone.lib="Porcine small intestine cDNA library"  
/note="directionally cloned cDNA in XLI-blue MRF."

BASE COUNT 80 a 65 c 34 g 101 t 11 others  
ORIGIN

Query Match 7.6%; Score 17; DB 23; Length 291;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 CACTCATCAACAATA 183  
|||||  
Db 178 CACTCATCAACAATA 194

RESULT 10  
H31005 352 bp mRNA EST 12-DEC-1996  
LOCUS SW3ICA272SR Brugia malayi infective larva cDNA (SAM94WL-Bml3)  
DEFINITION Brugia malayi cDNA clone SW3ICA272 5', mRNA sequence.  
ACCESSION H31005  
NID 9902161  
VERSION H31005.1 GI:902161  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Brugia malayi.  
Brugia malayi.  
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;  
Filarioidea; Onchocercidae; Brugia.  
REFERENCE  
1 (bases 1 to 352)  
Blaxter, M.L., Raghavan, N., Ghosh, I., Guiliano, D., Lu, W., Williams, S.A., Slatko, B. and Scott, A.L.  
Genes expressed in Brugia malayi infective third stage larvae  
Mol. Biochem. Parasitol. 77 (1), 77-94 (1996)  
On May 9, 1995 this sequence version replaced gi:802521.

CONTACT: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA  
Tel: 413/5853826  
Fax: 413/5853786  
Email: genome@smith.edu  
Similarity to hypothetical 35.8 KD protein of C.elegans  
Seq primer: pBluescript SK.

FEATURES  
source  
1. .352  
/organism="Brugia malayi"  
/strain="TRS Labs"  
/db\_xref="taxon:6279"  
/clone="SW3ICA272"  
/clone.lib="Brugia malayi infective larva cDNA (SAM94WL-Bml3)"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: lambda UniZap XR; Site.1: EcoR I; Site.2: Xho I; Lymphatic filarial nematode parasite of humans. mRNA was prepared from third stage infective larvae of Brugia malayi isolated from mosquitoes 10 days after infection and converted to double stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNase I. The library had 1.6 x 10<sup>6</sup> independent recombinants and average insert size was 900 base pairs. The library was constructed by Wenhong Lu. The library is available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 117 a 55 c 89 g 89 t 2 others  
ORIGIN

Query Match 7.6%; Score 17; DB 23; Length 352;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 ATTCAGCTGAACAACA 113  
|||||  
Db 120 ATTCAGCTGAACAACA 136

RESULT 11  
H43477 492 bp mRNA EST 31-JUL-1995  
LOCUS Y06406.r1 Soares breast 3NBH8t Homo sapiens cDNA clone  
DEFINITION IMAGE:182722 5', mRNA sequence.  
ACCESSION H43477

NID 9919529  
VERSION H43477.1 GI:919529  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 492)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfsing, T., Soares, M., Tan, F.,  
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P., and  
Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On Sep 21, 1992 this sequence version replaced gi:279074.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1241  
High quality sequence stops: 330  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1241 Std Error: 0.00  
Seq primer: M13Rpl  
High quality sequence stop: 330.  
Location/Qualifiers  
1..492  
/organism="Homo sapiens"  
/db\_xref="GDB:3815922"  
/db\_xref="taxon:9606"  
/clone="IMAGE:182722"  
/clone\_lib="Soares breast 3MDH8st"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pTZ19 (Pharmacia) with a  
modified polylinker. Site 1: Not I. Site 2: Eco RI. 1st  
strand cDNA was primed with a Not I - oligo(dT) primer (5'  
TGTTACCACTGAGTGGAGCGCCCTTTTCTTTTCTTTT 3'),  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of a modified pTZ19 vector (Pharmacia).  
Library went through one round of normalization to a cot =  
20. Library constructed by Bento Soares and M.Felina  
Bonafido."

BASE COUNT 154 a 83 c 97 g 153 t 5 others  
ORIGIN

Query Match 7.6%; Score 17; DB 23; Length 492;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ATTCTTCACCTCTAGA 74  
|||||  
DB 290 ATTCTTCACCTCTAGA 274

RESULT 12  
LOCUS D18308 344 bp mRNA EST 12-DEC-1995  
DEFINITION M08501184 Mouse 3'-directed Mus musculus cDNA clone md1421 3',  
mRNA sequence.  
ACCESSION D18308  
NID 91089632  
VERSION D18308.1 GI:1089632  
KEYWORDS EST.

SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 344)  
AUTHORS Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M., and Matsubara, K.  
TITLE Analysis of gene expression in mouse embryogenesis by 3'-directed  
cDNA sequencing  
JOURNAL Unpublished (1995)  
COMMENT On May 8, 1995 this sequence version replaced gi:799879.

CONTACT: Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M., and  
Matsubara, K.  
Institute for Cellular and Molecular Biology  
Osaka University  
3-1 Yamada-Oka, Suita, Osaka 565, Japan  
Insert Length: 1018 Std Error: 0.00  
High quality sequence stop: 248.  
Location/Qualifiers  
1..344  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="md1421"  
/clone\_lib="Mouse 3'-directed"  
/note="embryo"

BASE COUNT 78 a 76 c 80 g 105 t 5 others  
ORIGIN

Query Match 7.6%; Score 17; DB 24; Length 344;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 CTCCTATCTGTTGAA 222  
|||||  
DB 85 CTCCTATCTGTTGAA 101

RESULT 13  
LOCUS H74066/c 581 bp mRNA EST 31-OCT-1995  
DEFINITION Y513908 r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
IMAGE:214718 5', mRNA sequence.  
H74066  
ACCESSION H74066  
NID 91047202  
VERSION H74066.1 GI:1047202  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 581)  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonafido, M.F., Chiappelli, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfsing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
On May 8, 1995 this sequence version replaced gi:800218.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1291  
High quality sequence stops: 299

Source: IMAGE Consortium, LNL.  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1291 Std Error: 0.00  
Seq primer: M13Rp1  
High quality sequence stop: 299.  
Location/Qualifiers  
1.581  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:214718"  
/clone\_lib="Soares fetal liver spleen INF1S"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACGAGAGATTAATTAAGTCTTTTCTTTTCTTTTCTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 158 a 93 c 119 g 201 t 10 others  
ORIGIN

Query Match 7.6%; Score 17; DB 24; Length 581;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 CTGAGTAATTCATCAT 98  
|||||  
DB 23 CTGAGTAATTCATCAT 7

RESULT 14  
LOCUS N21299 580 bp mRNA EST 19-DEC-1995  
DEFINITION yx3je08.s1 Soares melanocyte 2NBM Homo sapiens cDNA clone  
IMAGE:265478 3', mRNA sequence.  
ACCESSION N21299  
NID 91126469  
VERSION N21299.1 GI:1126469  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 580)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
The WashU-Merck EST Project  
Unpublished (1995)  
On May 5, 1995 this sequence version replaced gi:798186.

TITLE  
JOURNAL  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
High quality sequence stops: 375  
Source: IMAGE Consortium, LNL.  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 641 Std Error: 0.00  
Seq primer: m13 -40 forward  
High quality sequence stop: 375.

FEATURES  
source Location/Qualifiers  
1.580  
/organism="Homo sapiens"  
/db\_xref="GDB:3875120"  
/db\_xref="taxon:9606"  
/map="11"  
/clone="IMAGE:265478"  
/clone\_lib="Soares melanocyte 2NBM"  
/issue\_type="melanocyte"  
/sex="male"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCTGAGTGGAGCGCGCCGACTTTTCTTTTCTTTTCTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library constructed by Bento Soares and  
M.Fatima Bonaldo. RNA from normal foreskin melanocytes  
(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 181 a 117 c 98 g 180 t 4 others  
ORIGIN

Query Match 7.6%; Score 17; DB 24; Length 580;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 TGAACACATGCGCTAT 120  
|||||  
DB 507 TGAACACATGCGCTAT 523

RESULT 15  
LOCUS N28457/c 537 bp mRNA EST 04-JAN-1996  
DEFINITION yx3jg05.r1 Soares melanocyte 2NBM Homo sapiens cDNA clone  
IMAGE:263768 5' similar to gb:M25393 T-CELL PROTEIN-TYROSIINE  
PHOSPHATASE (HUMAN);, mRNA sequence.  
ACCESSION N28457  
NID 91146693  
VERSION N28457.1 GI:1146693  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 537)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
The WashU-Merck EST Project  
Unpublished (1995)  
On Nov 22, 1995 this sequence version replaced gi:1071543.

TITLE  
JOURNAL  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
High quality sequence stops: 425  
Source: IMAGE Consortium, LNL.  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1915 Std Error: 0.00  
Seq primer: T7  
High quality sequence stop: 425.  
Location/Qualifiers  
1.537

```

/organism="Homo sapiens"
/db_xref="GDB:3873410"
/db_xref="taxon:9606"
/clone="IMAGE:263768"
/clone_lib="Soares melanocyte 2NDHM"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/note="vector: pT73b (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer [5',
TGTACCAATCTGAGTGGAGCGCCGACGTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."
BASE COUNT      132 a      140 c      138 g      112 t      15 others
ORIGIN

```

```

Query Match      7.6%; Score 17; DB 24; Length 537;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 203 CATCTGTGATCTGTT 219
   |||
DB 313 CATCTGTGATCTGTT 297

```

Search completed: August 6, 1999, 09:22:22  
Job time: 6213 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: August 6, 1999, 13:33:31 ; Search time 3075.15 Seconds

(without alignments)  
1804.379 Million cell updates/sec

Title: US-09-049-696-18

Perfect score: 2813

Sequence: 1 GAATTCACAGGAGATGTC.....AAATAATCATTCATCTTA 2813

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

ESR: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*  
45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	646.4	23.0	734	49	A1660957 w120d08.x
C 2	481.4	17.1	716	49	A1660234 we68902.x
C 3	473.6	16.8	480	35	AA581198 n038c07.x
C 4	471.2	16.8	501	51	A1721275 as82h08.x
C 5	347.8	12.4	378	31	AA296955 EST112726
C 6	336.8	12.0	646	37	AA691335 vs14h06.x
C 7	326	11.6	479	35	AA541829 vj01f01.x
C 8	322.2	11.5	335	51	A1721121 as73d08.x
C 9	309	11.0	428	30	284017 SS284017 Po
C 10	303.2	10.8	455	39	AA871197 vq32c11.x
C 11	298.2	10.6	450	37	AA734088 vs19c08.x
C 12	293.4	10.4	310	31	AA297150 EST112734
C 13	289.8	10.3	629	34	AA518112 v123c06.x
C 14	279	9.9	490	34	AA508854 n123h10.s
C 15	271.6	9.4	552	37	AA688953 vs04h07.x
C 16	263.4	9.4	507	36	AA623217 v114d07.x
C 17	261.8	9.3	554	36	AA623196 v114d07.x
C 18	259.4	9.2	469	36	AA611564 v085d05.x
C 19	232	8.2	255	20	D25727 HUMG504094
C 20	229.6	8.2	449	38	AA759816 vt81b01.x
C 21	227	8.1	482	37	AA711228 vt70c11.x
C 22	218.2	7.8	488	37	AA733415 vt75b06.x
C 23	213.8	7.6	392	50	A1687981 wa77f06.x
C 24	200	7.1	452	39	AA822200 vp36d05.x
C 25	198.4	7.1	638	45	A1390317 mx03b07.y
C 26	192.6	6.8	441	26	W41083 mc41c10.x1
C 27	190.6	6.8	301	37	AA689856 vf62c11.x
C 28	189.8	6.7	524	39	AA871630 vq39a03.x
C 29	171.6	6.1	370	37	AA711143 vt56d09.x
C 30	168.8	6.0	431	30	AA238284 my34g09.x
C 31	167.2	5.9	418	35	AA596289 vo26a08.x
C 32	163.2	5.8	327	23	F15082 SSC6A10 Por
C 33	162.8	5.8	463	33	AA839323 v040f07.x
C 34	162	5.8	297	49	A1641931 vq50e03.y
C 35	159	5.7	420	49	A1641868 vq39a03.y
C 36	158	5.6	470	43	A1197242 uc81f01.x
C 37	151.8	5.4	271	30	AA245498 mv51e11.x
C 38	149.4	5.3	427	34	AA498535 v040f12.x
C 39	149.4	5.3	421	37	AA691586 vs12g06.x
C 40	149.4	5.3	388	37	AA734161 vs19g06.x
C 41	144.4	5.1	297	40	AA929159 vq38g07.x
C 42	139.8	5.0	161	48	A1582072 at96a11.x
C 43	134.2	4.8	386	37	AA692521 vt59g08.x
C 44	133.8	4.8	468	49	A1640041 ms67c10.y
C 45	133.8	4.7	372	38	AA755896 v004f06.x

#### ALIGNMENTS

RESULT 1  
LOCUS A1660957 734 bp mRNA EST 10-MAY-1999  
DEFINITION w120d08.x1 Soares,Dieckgraefe-colon\_NHUC Homo sapiens CDNA clone  
IMAGE:2351151 3' similar to TR:088826 O88826 GOB-5 PROTEIN. ; mRNA  
ACCESSION A1660957  
NID g4764540





CHLORIDE CHANNEL PROTEIN ; mRNA sequence.  
 ACCESSION AI721275  
 MID 95038531  
 VERSION AI721275.1 GI:5038531  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eukaryota; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 501)  
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Matra,M.,  
 Martin,J., Moore,B., Schellenberg,K., Sepecoe,M., Tan,F.,  
 Thaising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 WashU-NCI human EST Project  
 TITLE Unpublished (1997)  
 JOURNAL On Jun 22, 1998 this sequence version replaced gi:3246918.  
 COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through INL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Possible reversed clone; similarity on wrong strand  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 394.  
 Location/Qualifiers

## FEATURES

SOURCE  
 1..501  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="Xp11.2"  
 /clone="IMAGE:2335263"  
 /clone\_lib="Barstead colon HPLNB7"  
 /sex="male"  
 /dev\_stage="adult, age 25"  
 /lab\_host="DH10B (phage resistant)"  
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTCAGATCGAGTCGAGCGGCCGCCCTTTTCTTTTCTTTTCTTTTCTTTT  
 3']; double stranded cDNA was ligated to Eco RI adaptors  
 [5' AATTCACGTAAT 3' and 5' ATTACTAGTG 3'], digested  
 with Not I and cloned into the Not I and Eco RI sites of  
 the modified pT73 vector. Library constructed by Bob  
 Barstead."  
 BASE COUNT 158 a 83 c 106 g 154 t  
 ORIGIN

Query Match 16.8%; Score 471.2; DB 51; Length 501;  
 Best Local Similarity 98.3%; Pred. No. 3e-121;  
 Matches 476; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 2329 ATTATCTGACTTGGACAGCTCTGGGATGATTTAGACATGGAACAGCTCACAAGTAT 2388  
 DB 501 ATTAATCTGACTTGACACTCTCTGGGATGATTTAGACATGGAACAGCTCACAAGTAT 442  
 QY 2389 ATCATTCGATAGTACGATGATTTCTTGTATCCAGAGCAAGTCAATGATCTCTCA 2448  
 DB 441 ATCATTCGATAGTACGATGATTTCTTGTATCCAGAGCAAGTCAATGATCTCTCA 382  
 QY 2449 GTGATATCTACTGCTCTCAATCCCAAGAGCAAGCACTGAGGAGACTCTTTTGTATAA 2508  
 DB 381 GTGATATCTACTGCTCTCAATCCCAAGAGCAAGCACTGAGGAGACTCTTTTGTATAA 322  
 QY 2509 CCAGAAACATTTACTTTGAAAAATGGCAGACAGATCTTTTATTGCTATTAGAGCTGTGAT 2568  
 DB 321 CCAGAAACATTTACTTTGAAAAATGGCAGACAGATCTTTTATTGCTATTAGAGCTGTGAT 262  
 QY 2569 AAGGTGATCTGAAATCAGAAATATCCAAATTCAGACGAGTATCTTTGTTATCTCTCA 2628

|||||  
 Db 261 AAGGTGATCTGAAATCAGAAATATCCAAATTCAGACGAGTATCTTTGTTATCTCTCA 202  
 QY 2629 CAGACTCCGCGAGAGACACCTAGTCTGATGAACGTCGTCTCTTCTTAATATTCAT 2688  
 Db 201 CAGACTCCGCGAGAGACACCTAGTCTGATGAACGTCGTCTCTTCTTAATATTCAT 142  
 QY 2689 ATCAACAGCCACCATCTCTGGCATTCACATTTTAAATATGATGAGAGTATGAGAGA 2748  
 Db 141 ATCAACAGCCACCATCTCTGGCATTCACATTTTAAATATGATGAGAGTATGAGAGA 82  
 QY 2749 CTGCAGCTGTAATAGCTAGGCTGATTTTGTTCAGATTAATAATAATCATTCAT 2808  
 Db 81 CTGCAGCTGTAATAGCTAGGCTGATTTTGTTCAGATTAATAATAATCATTCAT 22  
 QY 2809 CCTT 2812  
 Db 21 CCTT 18

RESULT 5  
 AA296955 378 bp mRNA EST 18-APR-1997  
 LOCUS AA296955  
 DEFINITION Colon I Homo sapiens cDNA 5' end, mRNA sequence.  
 ACCESSION AA296955  
 MID 91949515  
 VERSION AA296955.1 GI:1949515  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eukaryota; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
 1 (bases 1 to 378)  
 White,C.J., Lee,N.H., Kirkness,E.F., Weinstein,R.C., Man-Val,C.,  
 Blitt,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C.,  
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
 Fitzgerald,L.M., Fitzhugh,W.M., Frichman,J.L., Geoghagen,N.S.,  
 Glodex,A., Gnehm,C.L., Hanna,M.C., Hedlorn,E., Hinkle,P.S., Jr.,  
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,  
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
 Dilke,D., Feng,D.F., Ferrie,A., Fischer,C., Hastings,G.A.,  
 He,W.W., Hu,J.S., Greene,J.M., Gruber,U., Hudson,P., Kim,A.K.,  
 Kozak,D.L., Kunsch,C., Hungjun,D., Li,H., Weissner,P.S., Olsen,H.,  
 Raymond,L., Wel,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
 Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
 Fraser,C.M., and Venter,J.C.  
 TITLE Initial assessment of human gene diversity and expression patterns  
 based upon 85 million nucleotides of cDNA sequence  
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)  
 MEDLINE 96026280  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1288227.  
 Other ESTs: THC167738  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

## FEATURES

Source  
 1..378  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (Inhost):127834"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Colon I"



```

/dev_stage="adult"
/note="Organ: colon; Vector: pbluescript SK-; Site_1
ECORI, Site_2; XhoI"
BASE COUNT      112 a      79 c      64 g      119 t      4 others
ORIGIN

```

Query Match	12.4%;	Score 347.8;	DB 31;	Length 378;
Best Local Similarity	98.9%;	Pred. No. 9.5e-87;		
Matches 349;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;

[illegible]

2503 TTTAAACGAGAAACATTACTTTTGAATAATGGACAGACTTTTCAATTGCTATTCAGGCT 256  
|||||  
61 TTTAAACGAGAAACATTACTTTTGAATAATGGACAGACTTTTCAATTGCTATTCAGGCT 120

2563 GTTGATTAAGTCGATCTGAATCAGAAATATCCACACATTGCACGAGTATCTTTGTTTATT 2622

181 CCTCCACGACTCCGCCAGACACCTAGTCTCGATGAACGCTGCTCCTTCTCTAAT 240

241 ATTCATTCACACGACCATTCTCGCATTCACATTTTAAATATGTGGAAGTGATA 300

301 GGAGAACTGCAGCTGTCAATAGNCTAGGGCTGAATTTTGTCAAGGNNATAAA 353

RESULT	6
AA691335	
LOCUS	
AA691335	646 bp
	mRNA
EST	
16-DEC-1997	

clone IMAGE:1138235 5' similar to SW:ECCL\_BOVIN P54281 EPITHELIAL CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.  
AA691335

```

VERSION AA691335.1 GI:2692271
KEYWORDS EST.
SOURCE house mouse.

```

REFERENCE  
1 (bases 1 to 646)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

TITLE	The WashU-HMI Mouse EST Project
JOURNAL	unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1397564.

Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501 St Louis MO 63108

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [mousest@watson.wustl.edu](mailto:mousest@watson.wustl.edu)  
 This clone is available royalty-free through ILT: contact the

IMAGE CONSULTEUM (info@image.amaa.gov) for information.  
MGI:619507.  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stored: 491.

## FEATURES

```

source
1. .646
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/map="17q21; 2"
/clone="IMAGE:1138235"
/clone_1id="Barstead mouse irradiated colon MPLRB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: pUT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: EcoRI. Site_2: NotI. Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
[5'TGTATCAATCTGAAATGGGAGCGCCGCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors [AATTCGATCCGTC], digested with Not I and cloned
into the Not I and Eco RI sites of the modified pUT3
vector. Library constructed by Bob Barstead."
171 a 176 c 153 g 146 t

```

Query Match	Score	DB	Length
Best Local Similarity	74.48	Pred. No. 1.4e-83	
Query Match	12.08	Score 336.8	DB 37
Best Local Similarity	74.48	Pred. No. 1.4e-83	

2103 TGGAGGACCTGACATACCTGGCTGGATTGGAGATGATGAATACATGGAATCCACCAAG 2162

72	TCTGAAACAGT-----ATGTTAAGACAGCAGCTGCTT--CACGGACATCTC	123
2163	ACCTGAATTAATAAGATGATGTTCAACACCAAGTCGTGTTTCAGCAGAACATCTC	2222

124 AGGGGATCGTTGTGGCCCAATGCCCCCGAGCAGTGTCCATTCTGTACTCTTCC 183  
2223 GGGAGGCATATTGTGGCTCTGATGTC--CAATGCTCCATACCTGATCTCTTCC 22/9

2280 ACCGCGCCAAACACCGCAGCCTGGAGAGCGGAAAATTCACGCGGCCGACGCTCATATTATCTGC 2339

184 ACCCTGTCAATCACTAGCCTGGAGGCCAGCATCCGAAGGCGAACCCTGTGTAATCTAC 243

244 GTGACGGCTCTCTGGGGA TACTACGACCACGGAGAGCTTCCAATCATCCGAAT 303

304 GAGCACCACGATTCGTTGATCTCTCAGGACCCACTTCAACACCCTCACTCCCAAGTGAACACTAC 363

364 CGGTCTTATCCCCAAGAGGCCACGCTCTGAGGAAATCTTTGACTTTGCACTGGGAGGCCAA 423

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
424 CACTTTGGAAATGGCACAGATACTTCATTGCTATCCAGCGTGTGAAGAAGCACAACCT 483

|||||  
484 GAATCAGAATCTCCACATTTGCACGGGTGTCTGTGTTCAATCCCCGGCTAG----- 536

536 -GAGCGCCATTCCGAGACTCACTGCGCCCTTGCTCTGACATCAGCATCAACAGCAC 594

595 CATTCTGGCATCCACGCTGTAAGATAATGTGGAAGTGGCTAGGGAAATG 646

## RESULT 7

7

LOCUS	AA541829	479 bp	mRNA	EST	30-JUL-1997
DEFINITION	vJ01f01.r1 Barstead mouse pooled organs MFLRB4 Mus musculus CDNA clone IMAGE:320473 5' similar to TR:G1184066 G1184066				
ACCESSION	AA541829				
NID	92288263				
VERSION	AA541829.1	GI:2288263			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 479) Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.				
TITLE	The WashU-HMT Mouse EST Project				
JOURNAL	Unpublished (1996)				
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1406841.				
FEATURES	<p>source</p> <p>Contact: Marra M/Mouse EST Project WashU-HMT Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LNLN; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MG1:532689 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 368. Location/Qualifiers 1. 479 /organism="Mus musculus" /strain="FVB/N" /cd_xref="takon:10090" /clone_image="920473" /clone_lib="Barstead mouse pooled organs MFLRB4" /sex="mixed" /tissue_type="pooled organs" /dev_stage="7 day" /lab_host="DH10B" /note="Organ: pooled; Vector: pUT73D-Pac (Pharmacia) with a modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGATCTCAAGTGGAGGAGGCGCGCCCTTTTGTGTAATTAACGATTAAGGGA 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTGTTACGTCGATAC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pUT73 vector library constructed by Bob Barstead."</p>				
BASE COUNT	139 a	115 c	120 g	105 t	
ORIGIN					
Query Match	11.6%	Score 326;	DB 35;	Length 479;	
Best Local Similarity	80.98;	Pred. No. 1.3e-80;			
Matches 380; Conservative	0;	Mismatches 90;	Indels 0;	Gaps 0	
1676	CCTACCTCCCAATCCGACGATTTGGTGAAGTTGGCACTTGGAAATATACAGTCTGCAAGCAA	1735			
11					
5	CCAACTTCAAGTCCGACGACGCGTAAAGTTGGCTTTTGGAAATTAACGATTAAGGGA	64			
1736	GCTCACAACCTTGACCTGACTGTCACGTCCTGCGTCCCAATGCTACCTGCTCCAA	1795			
11					
65	GCTCACAAGACTCTCACTTGAAGTCTCACCTCCGTCGACGACAGTCTACACGCGCTCTTA	124			
1796	TTACAGTACTTCCAAAGCAAGCAAGGACACGAGAAATTTCCCGACGCTTGAGATT	1855			
11					
125	TTTACAGTACCCCGGTAGTGAATTAAGAACACACAGGGAATTTCCCGACGCTTGTAACAGTGT	184			

```

OY 1856 ATCGAATATTTCGCCAGAGAGCCCTCCCAATTCACAGGCGCAGTGCACAGCCCTGATG 1915
      |||||
      |ATGCAAGCTATTCGCCAGAGAGCCCTCCCAATTCACAGGCGCAGGCTCACAGCCTTGATG 244
OY 1916 AATCAGTGAATGAAAAACAGTACTCTTGGAACTACTGGATATATGAGCAGAGTCTGATG 1975
      |||||
      |AATCTGTGAATGAAAAACAGTAACTCCGTGAATTAAGTGAATTAAGCAGAGTCCGATG 304
OY 1976 CTACTAAGATGAGCGGTCTCTACTTCAGAGTATTTCACAACTTAAGACACGAAATGGTATG 2035
      |||||
      |CCACCAAAAGAAAGATGCTCTACATCAAGGTTTTTTTACAGCTTTTGAATGCAATGTTAT 364
OY 2036 ACAGTGTAAAGTGCGGCGCTCTGGAGAGAGTAAACGACGCGACAGCAGAGTGTATACCC 2095
      |||||
      |ACAGCGTTAAATATATGCGCTCTGGAGAGAGTCACTCTACAGACAGACAGACAGACCTC 424
OY 2096 AGCAGATGAGCAGCATGTACATACCTGCGCTGGATTTAGACATGATGAATA 2145
      |||||
      |CGAAGAACAGAGCCATGTACATAGATGCGCTGGATTTAGAGATGATGTAAGTA 474
Db 425

RESULT 8
A1721121/c 335 bp mRNA EST 10-JUN-1999
LOCUS as73d08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
DEFINITION IMAGE:23134351 3' similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA
sequence.
A1721121
A1721121.1 GI:5038377
EST.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 335)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S.,
Kizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Matra,M.,
Martin,J., Moore,B., Schellenberg,K., Stepien,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
On Jun 22, 1998 this sequence version replaced gi:3246762.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco
High quality sequence stop: 312.

FEATURES
source
Location/Qualifiers
1..335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:23134351"
/clone_1ib="Barstead colon HPLRB7"
/sex="male"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5
TGTTCGATCTGATCGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[5' AATTCAGTAGAT 3' and 5' ATTACAGAG 3'], digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pT73 vector. Library constructed by Bob

```

	Barstead."
BASE COUNT	98 a 57 c 71 g 109 t
ORIGIN	

Query Match	11.5%	Score 322.2;	DB 51;	Length 335;
Best Local Similarity	97.6%;	Pred. No. 1.3e-79;		
Matches 327;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0

Qy	2416	GAAGCCAACTGTGAGGAGAGTCTTTTGGTTAAACCGAAACCTTACTTTGAAATGGG	2535
Db	335	GAAGCCAACTGTGAGGAGAGTCTTTTGGTTAAACCGAAACCAATTACTTTGAAATGGG	276
Qy	2536	ACAGATCTTTTATGTCTATTCAGAGGTGTGTAAAGTGTGATGTGAAATCAGAAATATTC	2595
Db	275	ACAGATCTTTTATGTCTATTCAGAGGTGTGTAAAGTGTGATGTGAAATCAGAAATATTC	216
Qy	2596	AACATTGCACGAGTATCTTTTGTATTCTCCACAGACTCCGACAGAGACACTAGTCTCT	2655
Db	215	AACATTGCACGAGTATCTTTTGTATTCTCCACAGACTCCGACAGAGACACTAGTCTCT	156
Qy	2656	GATGAAGCTGCGTCCCTTCCTGCTAATTCATATCAATCAAGACCACTTCCTGGCATTCAC	2715
Db	155	GATGAAGCTGCGTCCCTTCCTGCTAATTCATATCAATCAAGACCACTTCCTGGCATTCAC	96
Qy	2716	ATTTTAAAAAATATGTGGAAGTGATAGGAGAACTGCAGCTGTCAATAGCCTAGGGCTGA	2775
Db	95	ATTTTAAAAAATATGTGGAAGTGATAGGAGAACTGCAGCTGTCAATAGCCTAGGGGAAAA	36
Qy	2776	ATTTTGTCAAGTAATAATAATTAATCAATTCATCC 2810	
Db	35	ATTCCTATCAGATAATAATAATAATCAATTCATCC 1	

RESULT	9
LOCUS	284017
DEFINITION	284017 428 bp mRNA EST 30-JAN-1997
ACCESSION	SS284017 Porcine small intestine cDNA library Sus scrofa cDNA clone
NID	c12a03.5', mRNA sequence.
VERSION	284017
KEYWORDS	g1806336
SOURCE	284017.1 GI:1806336
ORGANISM	EST.
REFERENCE	p19.
AUTHORS	Sus scrofa
TITLE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata; Mammalia;
JOURNAL	Eutheria; Cetartiodactyla; Suidae; Sus.
MEDLINE	1 (bases 1 to 428)
COMMENT	Wintero,A.K., Fredholm,M. and Davies,W.
	Evaluation and characterization of a porcine small intestine cDNA
	library: analysis of 839 clones
	Mamm. Genome 7 (7), 509-517 (1996)
	96327607
	On Sep 12, 1996 this sequence version replaced gi:1404772.

**CONTACT:** A. K. Wintereoe  
Department of Animal Science and Animal Health, Division of Animal  
Genetics, The Royal Veterinary and Agricultural University  
Bulowsvej 13, 1870 Frederiksberg C, Denmark.  
**LOCATION/QUALIFIERS**  
1. .428

BASE COUNT	134 a	87 c	100 g	102 t	5 others
ORIGIN					
Query Match	11.0%	Score 309;	DB 30;	Length 428;	
Best Local Similarity	82.4%;	Pred. No. 7.2e-76;			
Matches 351;	Conservative	0;	Mismatches 75;	Indels 0;	Gaps 0

Oy	2	AAATCACAGGGAGATGTACACCAATGGGGCCATTAGAGCTTCGTGTATCTTGAATTC	61
Db	3	AAACCAAGGGAGATGTACACCAATGGAGTTCATTAGAGTTCGTGTATCTTGAATTC	62
Oy	62	TTACACTCTTGAAGAGGGGCCCTGAGTAATTCATCTATTCACGTGAACAAATGGCTATG	121
Db	63	TTACACTCTCTGAAGAGGGGCTTGAGTAATTCCTGCTATCCAGCTGAAGGCAATGGGCTACG	122
Oy	122	AAGGCAATGTGGTTCGAATGCACCCCAATGGCCAGAGATGAAACCTCATTCACAA	181
Db	123	AGGGCAATGTATTCGAATTAACCCCAAGCTGCCAGAGATGAAACCTCATTCAGATA	182
Oy	182	TAAAGGACATGGTAGCCAGGCACTCTGTATCTGTATTGTTGAAGCTACAGAAAGCATTTT	241
Db	183	TAAAGGATATGGTAGCCAAACATCTCCATATCTGTTGAAGCTACAGNAAAAAGATTCT	242
Oy	242	ATTTCAAAAAGTTGCCATTTTGATTTCTTCTGAACATCGAAGACAAAGGCTGACTATGTGA	301
Db	243	ATTTCAAAAAGTTGCCATTTTGATTTCTTCTGAACATCGAAGGACAAAGGCTGACTATGTGA	302
Oy	302	GACCAAAACTTGAGACCTACAAAAATGCTGATGTTCTGTGTGCTGAGTCTACTCTCCAG	361
Db	303	AAACCAAACTTGAGACCTACAAAAATGCTGATGTTGTGTGTACTGAACCAAAATCTCCAG	362
Oy	362	GTAATGATGAACCTACACTGAGCAGATGGGCACTGTGAGAGAGAGGGTGAAGGATCC	421
Db	363	AAAATGATGGGCCCTATACAGAGCAGATGGGNAACNTGTGAGAAAAGGGTGAAGAAATTT	422
Oy	422	AACCTCA 427	
Db	423	ATTTC A 428	

RESULT 10  
AA871197  
LOCUS  
DEFINITION  
AA871197 455 bp mRNA EST 16-MAR-1996  
v932c11.r1 Barstead bowel MPELB9 Mus musculus cDNA clone  
IMAGE:1095956.5' similar to SW:EDIC.BOVIN P54281 EPTHELINL  
CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.  
AA871197  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (Bases 1 to 455)  
Merrill, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuc  
Schell, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, R., Steptoe, M., Tan, F., Underwood, R., Moore, B.,  
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Sep 12, 1996 this sequence version replaced gi:1395394.

**FEATURES**  
**source**  
 Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.jnl.gov) for further information.  
 MGI:602188  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 414.  
 location/Qualifiers  
 1..455  
 /organism="Mus musculus"

```

/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1095956"
/clone_lib="Barstead bowel MPLRB9"
/tissue_type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACGATCTGAAGGAGGAGCGCCGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[ATTCGATCTTGG], digested with Not I and Eco RI sites of the
Not I and Eco RI sites of the modified pT73 vector.
Source irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob
Barstead."
BASE COUNT      122 a      123 c      105 g      105 t
ORIGIN

```

```

Query Match      10.8%; Score 303.2; DB 39; Length 455;
Best Local Similarity 79.4%; Pred. No. 3.1e-74;
Matches 359; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```

```

QY 1416 GAACAATGGCTCATGATGCTTTGGGGCCCTTCATCAGAAATGAGCTGTCTCA 1475
      ||||||| || ||||||| || || ||||||| || || ||||
Db 4 GAACAATGGCTGTGATGCTTTGCGACACTCTCAGGAATGCGGATCGCTCA 63

QY 1476 GCGCTCCATCCAGCTTGAGAGTAAGGANTTACCTCCAGAACCCAGTGAATG 1535
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 64 GCATCTCCATCCAGCTTGAGAGTAAGGANTTATCCAGTAATCAATGGATGATG 123

QY 1536 CACAGTGTGCGAGACACCGTGGGAAGACATTTGTTTATCACCCTGGACAAC 1595
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 124 CTCAGTATCGTGGACAGCTGCGTGGCAAGACACTTGTCTTATCACCCTGGACAAC 183

QY 1596 GCAGCTCCCAATCTCTCTGCGATCCAGTGACAGAACAGAGTGGCTTTAGT 1655
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 184 GCATCTCTCCATCAATTTATCTGGATCCCGAGCGAGGAAACAAATGTTTACT 243

QY 1656 GGACAAAACACCAAAATGGCTTCCCAATCCAGGATGCTAAGTGGAGACTG 1715
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 244 AGACCAACCACTAAGTGGCTTACCTCCAGGACGCTAAGTGGCTTTTG 303

QY 1716 GAATATCAGTGTGACAGACGCTGACCAACCTGACTGTACGTCCTCCGCTC 1775
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304 GAATATCAGCATTTCAAGGAGCTCAGACGTCTGACCTTGCACCTCCGCTGACG 363

QY 1776 CAATGCTACCTGCTCCATTTACAGTACTTCAAAAAGAACAGACAGCAAT 1835
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 364 AAGTCTTACGCTGCTCTTATACAGTACCCCGGTAGATAGAACAACAGGAAAT 423

QY 1836 CCCAGCCCTGTGAGTATTTAGCAATATTC 1867
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 424 CCCAGCCCTGTACAGTGTATGACAGCAATTC 455

```

```

RESULT 11
LOCUS      AA734088      450 bp      mRNA      EST      07-JAN-1998
DEFINITION vs19c08.t1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone IMAGE:1138670 5' similar to SW:ECIC, Bovin P54281 EPITHELIAL
            CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.
ACCESSION  AA734088
VERSION    92755755
KEYWORDS   AA734088.1 GI:2755755
SOURCE     EST.
ORGANISM   house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 450)

```

**AUTHORS**  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisler, S., Kucaba, T., Lacy, M., Martin, J., Morris, M.,  
 Schellberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 The WashU-HMNI Mouse EST Project  
 Unpublished (1996)  
 On May 8, 1995 this sequence version replaced gi:801722.

**TITLE**  
 JOURNAL  
 COMMENT  
 Contact: Marra M/Mouse EST Project  
 WashU-HMNI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:619942

**FEATURES**  
 source  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -28ml3 rev2 EX from Amersham  
 High quality sequence stop: 444.  
 Location/Qualifiers  
 1..450

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1138670"
/clone_lib="Barstead mouse irradiated colon MPLRB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
[5'TGTACCAATCTGAAGTGGAGGCGCCGCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors [ATTCGATCTTGG], digested with Not I and Eco RI
into the Not I and Eco RI sites of the modified pT73
vector. Library constructed by Bob Barstead."
BASE COUNT      132 a      109 c      112 g      97 t
ORIGIN

```

```

Query Match      10.6%; Score 298.2; DB 37; Length 450;
Best Local Similarity 80.0%; Pred. No. 7.8e-73;
Matches 351; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

QY 1731 AGCAAGCTCACAACCTTTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1790
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 12 AGCGAGCTCACAAGCTTCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 71

QY 1791 TCCATTTACGCTGCTTCCCAACAGACAGACAGCAATTTCCCGACCTGTG 1850
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 72 TCCATTTACGCTGCTTCCCAACAGACAGACAGCAATTTCCCGACCTGTG 131

QY 1851 AGTTATGCAATATATGCGCAAGAGGCTCCCAATTTCTCAGGCGCAGTGTACAGCC 1910
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 132 AGTGTATGCAAGATGCGCAAGAGGCTCCCAATTTCTCAGGCGCAGTGTACAGCC 191

QY 1911 GATTGATCACTGATGATGAAAAACAGTACTTGGAACTATGATATGAGACAGTGC 1970
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 192 GATTGATCACTGATGATGAAAAACAGTACTTGGAACTATGATATGAGACAGTGC 251

QY 1971 TGATGCTACTAAGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2030
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 252 CGATGCCACCAAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311

QY 2031 TAGATACAGTGTAAAGTGGCGCTGCGAGAGTAAACGACGACGAGAGAGTGTAT 2090
      ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 312 TAGATACAGGCTTAAATATATGGCTGCGAGAGTCACTTCAACAGACAGAGAGCAGC 371

```







This Page Blank (uspto)



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:32:54 ; Search time 3075.15 Seconds

(without alignments)  
161.644 Million cell updates/sec

Title: US-09-049-696-6

Perfect score: 252  
Sequence: 1 CAAGAAATGCTGTTTACT.....GAGGACCTCCATCTGCAGC 252

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est23: \*  
55: em\_est24: \*  
56: em\_est25: \*  
57: em\_est26: \*

Print. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	140.4	55.7	327	23	F15082	F15082 SSC6A10 Por
2	99.2	39.4	372	38	AA755896	AA755896 v04f06.r
3	86.8	34.4	480	35	AA581198	AA581198 nd38c07.r
4	33	13.1	338	21	R01958	R01958 y855c07.r1
5	31.6	12.5	354	32	AA354486	AA354486 EST62756
6	31	12.3	375	23	H58328	H58328 yf25602.r1
7	31	12.3	421	36	AA615628	AA615628 v071d08.r
8	31	12.3	442	40	AA929419	AA929419 my54908.r
9	31	12.3	208	44	AT241971	AT241971 qv51a03.x
10	30.8	12.2	183	43	AI223520	AI223520 qx27f06.x
11	30.6	12.1	423	35	AA356290	AA356290 145 Ldblo
12	30.6	12.1	539	45	AA859224	AA859224 UI-R-E0-C
13	30.6	12.1	442	48	AI599870	AI599870 EST251573
14	30.6	12.1	324	48	AI602041	AI602041 UI-R-G0-u
15	30.6	12.1	299	51	AT705570	AT705570 UI-R-AE1-
16	30.4	12.1	513	33	AA386850	AA386850 vc21f02.r
17	30.4	12.1	447	48	AI396884	AI396884 v08h08.y
18	30.2	12.0	427	39	AA856543	AA856543 oe47b10.s
19	29.6	11.7	557	35	AA538011	AA538011 vj34h06.r
20	29.6	11.7	693	42	AI098928	AI098928 ue37h02.y
21	29.6	11.7	490	42	AU023771	AU023771 AU023771
22	29.4	11.7	416	35	AA596974	AA596974 v016908.r
23	29.4	11.7	527	37	AA690789	AA690789 vt25902.r
24	29.4	11.7	481	37	AA711316	AA711316 vu25a09.x
25	29.4	11.7	438	45	AI376111	AI376111 ta04d09.x
26	29.2	11.6	118	41	AI007805	AI007805 EST202256
27	29.2	11.6	538	51	AT722806	AT722806 fc31h03.y
28	29	11.5	567	26	W52444	W52444 zd13f07.s1
29	29	11.5	474	26	W73692	W73692 zd30c01.s1
30	29	11.5	429	26	W84875	W84875 zh53910.r1
31	29	11.5	440	26	W95650	W95650 ze06h11.r1
32	29	11.5	480	28	AA119604	AA119604 mp88e10.r
33	29	11.5	568	30	AA203545	AA203545 zx59a05.r
34	29	11.5	260	34	AA512627	AA512627 vj19d08.r
35	29	11.5	555	36	AA610122	AA610122 at19h06.s
36	29	11.5	550	39	C88279	C88279 C88279 Mous
37	29	11.5	635	44	AA914718	AA914718 v203f01.r
38	29	11.5	635	44	AI316514	AI316514 u14907.x
39	28.8	11.4	459	34	AA467215	AA467215 v08h08.r
40	28.8	11.4	294	37	AA686617	AA686617 EST110825
41	28.8	11.4	449	48	AI594861	AI594861 v08h08.x
42	28.8	11.4	377	49	AV012003	AV012003 AVO12003
43	28.6	11.3	393	23	L11850	L11850 MUSEST34A.M
44	28.6	11.3	414	46	AI406850	AI406850 EST235137
45	28.6	11.3	396	48	AI578769	AI578769 UI-R-Y0-v

#### ALIGNMENTS

RESULT 1  
LOCUS F15082 327 bp mRNA  
DEFINITION SSC6A10 Porcine small intestine cDNA library Sus scrofa cDNA clone  
ACCESSION F15082  
NID 9972063  
VERSION F15082.1 GI:972063

KEYWORDS	EST.
SOURCE	pig.
ORGANISM	Sus scrofa
REFERENCE	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
TITLE	1 (bases 1 to 327)
JOURNAL	Wintero,A.K., Fredholm,M. and Davies,W.
MEDLINE	Evaluation and characterization of a porcine small intestine cDNA
COMMENT	library: analysis of 839 clones Mamm. Genome 7 (7), 509-517 (1996) 96327607 On May 8, 1995 this sequence version replaced gi:799705.
FEATURES	Source
BASE COUNT	78 a 82 c 91 g 72 t 4 others
ORIGIN	
Query Match	55.7%; Score 140.4; DB 23; Length 327;
Best Local Similarity	76.9%; Pred. No. 2.6e-36;
Matches 183; Conservative	0; Mismatches 53; Indels 2; Gaps 1;
4 AGAATGTGTGTTGTGTCCTTGACAAATCTGGAACATGAGCGACTGTACCGCCTCAAT 63	
1 AGAATCGTGTCTAGTCTCTCGACAAAGTCTGGAGACATGACCGTGTGGTGGCGCTTAGG 60	
64 CGACTGAATCAAGCAGCGCCAGCTTTTCTCTGCGCAGACAGTTCAGCTGGGGTCTCGGTT 123	
61 CGAGCATCAAGCAGCAGGCAAACTCTTCCTTCGACAGACAGTGGACAGAGGGCGTGGTT 120	
124 GGGATGTGACATTTGACAGTCTGCCCATGTACAAAGTGAATCATACAGATTAACAGT 183	
121 GGGATGTGGCGCTTTGACAGCGCTCTATGTAAAGAGTAAGTCTGCAGATTAACAGC 180	
184 G-GGAGTGACAGGACACACCTCGCCAAAGTTACTCGACACACTTCAGAGAGGAGC 239	
181 GCCGCGGATMGACACAGTCTTACCCGACGCTTACCCACACAGCAGCTCGGTGGAC 238	
RESULT 2	
AA755896	372 bp mRNA EST 21-JAN-1998
LOCUS	V004f06.r1 Soares mouse mammary gland RNMKG Mus musculus cDNA clone
DEFINITION	IMAGE:1179683 5' similar to SW:ELIC_BOVIN P54281 EPITHELIAL
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.	
AA755896	
NID	92803094
VERSION	AA755896.1 GI:2803094
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
1 (bases 1 to 372)	
Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,	
Geisel,S., Kucada,T., Lacy,M., Le,M., Martin,U., Morris,M.,	
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,	
Therising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and	
Waterson,R.	
The Mashu-HMI Mouse EST Project	
Unpublished (1996)	
On Jan 17, 1998 this sequence version replaced gi:2044941.	

Contact: Maria M/Mouse Est Project  
Mashu-HMI Mouse Est Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (infoimage.lnl.gov) for further information.  
MG:637531

Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 365.

```

FEATURES
Source
1..372
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map=x; 2:726H5:12"
/clone="IMAGE:1176683"
/clone_id="Soares mouse mammary gland NBMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTTCAATCTGTGAAGGGAGCGCCGCAGATGTTTTTTTTTTTTTTTTTTT
T 3'] double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldi."
BASE COUNT      121 a      86 c      81 g      84 t
ORIGIN
Query Match          39.4%; Score 99.2; DB 38; Length 372;
Best Local Similarity 63.3%; Pred.No.1.le-22;
Matches 152; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
OY   5 GAATGTGTGTTTAGTCCTTGACAAATGTGAGAAGATGCCAGTGTAAACGCCCTCATC 64
    ||||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   128 GAGTGTGTGTTGTTGTGAAATAAATTGGAGACAAGCAAACCGCTTATTC 187
OY   65 GACTGAATCAAGCAGGCGCAGCTTTTCCCTGCTGACAGACAGTTGAGCTGGGGTCTG 124
    ||||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   188 GAATGAATCAAGCAGCAACAAGCTGTAACCTAAATGTGAAAAGAGCTGTAGTTG 247
OY   125 GGATGTGTACATTGGACAGTGTGCTGCCATGTGCAAAATGGAAGTCAATACAGTAAC 184
    ||||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   248 GATTAGTACATTGGACAGCGCTGCCCATCCATCCAAAAATTAATTAATAAATTAAG 307
OY   185 GCAGAGACAGGAGACACATCGCCAAAAAATTAACCGCAGCAGCTCAGAGAGAGCTCCA 244
    ||||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   308 GTAGTGACTAACCAAGATCACCGCAACCCTCCCCCAACAGAGCTTCTGGGTGACACT 367
RESULT           3
LOCUS            AA581198
DEFINITION       AA581198 mRNA EST 05-JAN-1998
                  n838c07.r1 NCI_CGAP_COI Homo sapiens cDNA clone IMAGE:802572 5'
                  similar to TR:GI184066 GI184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.
                  ;, mRNA sequence.
ACCESSION        AA581198
NID              92358970
VERSION          AA581198.1 GI:2358970
KEYWORDS         EST.
SOURCE           human.
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```





TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Nov 29, 1993 this sequence version replaced g1:430449.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:386919  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 406.

## FEATURES

source 1..421

/organism="Mus musculus"  
/strain="C3H"  
/db\_xref="taxon:10090"  
/map="11"  
/clone.lib="IMAGE:1064559"  
/cell\_line="C2C12"  
/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCGATCTGAGTGGAGCGGCCCTTTTCTTCTGCTCCGGGATCATGCTGGG  
3'] double-stranded cDNA was ligated to Eco RI adaptors  
[AATTCGATCTCTG], digested with Not I and cloned into the  
Not I and Eco RI sites of the modified pT7T3 vector.  
Library constructed by Bob Barstead. The C2C12 cell line  
[available from ATCC, catalog # CRL-1772] differentiates  
rapidly, forming contractile myotubes and producing  
characteristic muscle proteins."

BASE COUNT 94 a 132 c 118 g 77 t  
ORIGIN

Query Match 12.3%; Score 31; DB 36; Length 421;

Best Local Similarity 56.3%; Pred. No. 4.2; Mismatches 45; Indels 0; Gaps 0;

Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 56 GCCTCATGACATGATCAAGCAGCCAGCTTCTCTGTCGACAGATGAGCTGGGT 115

DB 155 GCCACATCGCGTGCAGCATGTCAGACAGCTCTTCTGCTCCGGGATCATGCTGGG 96

QY 116 CCGGTGGGATGTCACATTTGACAGTCTGCCCATGTACA 158

DB 95 TATGGGTGGGATGTCGCGGTGTCGAGTCTTCTCTGCGACA 53

## RESULT 8

AA29419/c

LOCUS AA929419 442 bp mRNA EST 23-APR-1998

DEFINITION my54g08.r1 Barstead mouse pooled organs MRLB4 Mus musculus cDNA  
clone IMAGE:699710 5' similar to gb:M1560 FRUCOSE-BISPHOSPHATE  
ALDOLASE A (HUMAN); gb:Y00516 Mouse mRNA for aldolase A (MOUSE);

MRNA sequence.

ACCESSION AA929419

NID 93078728

VERSION AA929419.1 GI:3078728

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 442) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Tetelsting,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
COMMENT On Jan 19, 1998 this sequence version replaced g1:2153448.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:433270  
Trace considered overall poor quality  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 1.

## FEATURES

source 1..442

/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone.lib="IMAGE:699710"  
/clone.lib="Barstead mouse pooled organs MRLB4"  
/sex="mixed"  
/tissue\_type="pooled organs"  
/dev\_stage="7 day"  
/lab\_host="DH10B"

/note="organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTCGATCTGAGTGGAGCGGCCCTTTTCTTCTGCTCCGGGATCATGCTGGG  
3'] double-stranded cDNA was ligated to Eco RI adaptors  
[GTTCGATCTCTG], digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
Library constructed by Bob Barstead."

BASE COUNT 91 a 135 c 136 g 80 t  
ORIGIN

Query Match 12.3%; Score 31; DB 40; Length 442;

Best Local Similarity 56.3%; Pred. No. 4.2; Mismatches 45; Indels 0; Gaps 0;

Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 56 GCCTCATGACATGATCAAGCAGCCAGCTTCTCTGTCGACAGATGAGCTGGGT 115

DB 148 GCCACATCGCGTGCAGCATGTCAGACAGCTCTTCTGCTCCGGGATCATGCTGGG 89

QY 116 CCGGTGGGATGTCACATTTGACAGTCTGCCCATGTACA 158

DB 88 TATGGGTGGGATGTCGCGGTGTCGAGTCTTCTCTGCGACA 46

## RESULT 9

A1241971

LOCUS A1241971 208 bp mRNA EST 02-DEC-1998

DEFINITION qv51a03.x1 NCI-CGAP\_Ov32 Homo sapiens cDNA clone IMAGE:1985068 3'  
similar to gb:M12625 PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE  
(HUMAN); mRNA sequence.

ACCESSION A1241971

NID 93837368

VERSION A1241971.1 GI:3837368

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 208) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index



Query Match 12.1%; Score 30.6; DB 35; Length 423;  
 Best Local Similarity 47.0%; Pred. No. 5.7; Mismatches 98; Indels 0; Gaps 0;  
 Matches 87; Conservative 0;

23 TTGACAAATCTGAGACATGCGACCTGTAACCGCTCAATGCAATCAAGCAGCC 82  
 51 TGGACAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 110  
 83 AGCTTTCTGCTGACAGACAGTTGAGCTGGGCTCGGCTGGAGAGTGACATTTGACA 142  
 111 TGGCTTGTCTTTGATTAATTTGGCGGCGCTCTCTGCGGCTGACAGTGAATTTGGCAT 170  
 143 GTCTCTGCGCTGACAGAGAGTCAATCAATCAATCAATCAATCAATCAATCAAT 202  
 171 TCGCGCGCGGAGATTTTGGAGCAGAGATCAATCAATCAATCAATCAATCAATCAAT 230  
 203 TCGCC 207  
 231 TCCCC 235

# RESULT 12

AA859224 539 bp mRNA EST 05-FEB-1999  
 LOCUS AA859224  
 DEFINITION UI-R-E0-cd-b-09-0-UI s1 UI-R-E0 Rattus norvegicus cDNA clone  
 UI-R-E0-cd-b-09-0-UI 3' similar to dbj|D12771|RATNT2 Rat mRNA for  
 mitochondrial adenine nucleotide translocator, mRNA sequence.

ACCESSION AA859224  
 NID 94229581  
 VERSION AA859224.1 GI:4229581  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 539)  
 REFERENCE 1 (Bases 1 to 539)  
 AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2948575.

CONTACT: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: mscares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the  
 oligo-dT track served to identify it as a clone from the normalized  
 adult 8-day-embryo library. cDNA library Preparation: M. Fatima  
 Bonaldo, Ph.D. Clone distribution: clones will be available through  
 Research Genetics  
 Seq primer: M13 Forward  
 Location/Qualifiers  
 1..539  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-E0-cd-b-09-0-UI"  
 /clone\_11b="UI-R-E0"  
 /dev\_stage="embryonic"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT733-Pac (Pharmacia) with a modified  
 polylinker. Site.1: NotI; Site.2: EcoRI; This library  
 consists of a mixture of individually tagged normalized  
 libraries constructed from 8, 12 and 18-day embryo. The  
 tag is a string of 3-5 nucleotides present between the  
 Not I site and the oligo-dT track which allows  
 identification of the library of origin of a clone within  
 the mixture."

## FEATURES

source  
 1..539  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-E0-cd-b-09-0-UI"  
 /clone\_11b="UI-R-E0"  
 /dev\_stage="embryonic"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT733-Pac (Pharmacia) with a modified  
 polylinker. Site.1: NotI; Site.2: EcoRI; This library  
 consists of a mixture of individually tagged normalized  
 libraries constructed from 8, 12 and 18-day embryo. The  
 tag is a string of 3-5 nucleotides present between the  
 Not I site and the oligo-dT track which allows  
 identification of the library of origin of a clone within  
 the mixture."

BASE COUNT 151 a 101 c 148 g 139 t  
 ORIGIN

Query Match 12.1%; Score 30.6; DB 45; Length 539;  
 Best Local Similarity 62.3%; Pred. No. 6.3; Mismatches 29; Indels 0; Gaps 0;  
 Matches 48; Conservative 0;

13 TGTATGCTCTTGACAAATCTGAGACATGCGACCTGTAACCGCTCAATGACTGAAT 72  
 28 TCTTATGCTCTTATATACGTGTGAGCCCAAGCCACTGTATACAGAGAAATCAGTAGA 87  
 73 CAGCAGCCAGCTTT 89  
 88 CAGGGAAGCAGCGTT 104

# RESULT 13

A1599870 442 bp mRNA EST 21-APR-1999  
 LOCUS A1599870  
 DEFINITION EST251573 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone  
 REMER10 3' end, mRNA sequence.

ACCESSION A1599870  
 NID 94608918  
 VERSION A1599870.1 GI:4608918  
 KEYWORDS EST.  
 SOURCE Rattus sp.  
 ORGANISM Rattus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 442)  
 REFERENCE 1 (Bases 1 to 442)  
 AUTHORS Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J.,  
 Kerlavage, A.R. and Adams, M.D.  
 TITLE Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat  
 Gene Index  
 JOURNAL Unpublished (1998)  
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2948723.

CONTACT: Lee, NH  
 ATCC  
 The Institute for Genomic Research  
 9712, Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208  
 Email: nhlee@igf.org  
 Seq primer: M13-21.  
 Location/Qualifiers  
 1..442  
 /organism="Rattus sp."  
 /db\_xref="taxon:10116"  
 /clone="REMER10"  
 /clone\_11b="Normalized rat embryo, Bento Soares"  
 /dev\_stage="embryo 8, 12, 18 dpc"  
 /note="Vector: pT733Pac; Site.1: EcoRI; Site.2: NotI"

## FEATURES

source  
 1..442  
 /organism="Rattus sp."  
 /db\_xref="taxon:10116"  
 /clone="REMER10"  
 /clone\_11b="Normalized rat embryo, Bento Soares"  
 /dev\_stage="embryo 8, 12, 18 dpc"  
 /note="Vector: pT733Pac; Site.1: EcoRI; Site.2: NotI"

## BASE COUNT

131 a 81 c 125 g 105 t

Query Match 12.1%; Score 30.6; DB 48; Length 442;  
 Best Local Similarity 62.3%; Pred. No. 5.8; Mismatches 29; Indels 0; Gaps 0;  
 Matches 48; Conservative 0;

13 TGTATGCTCTTGACAAATCTGAGACATGCGACCTGTAACCGCTCAATGACTGAAT 72  
 11 TCTTATGCTCTTATATACGTGTGAGCCCAAGCCACTGTATACAGAGAAATCAGTAGA 70  
 73 CAGCAGCCAGCTTT 89  
 71 CAGGGAAGCAGCGTT 87

# RESULT 14

A1602041





Mon Aug 9 13:26:20 1999

us-09-049-696-6.rst

Page 9

Search completed: August 6, 1999, 13:32:58  
Job time: 12976 sec

---

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:42:14 ; Search time 650.92 Seconds

(without alignments)  
1146.566 Million cell updates/sec

Title: US-09-049-696-20

Perfect score: 2983  
Sequence: 1 GAAATCAGCAGGAGATGTCAC.....AATGCTAACAACACTGGGTA 2983

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685	23.0	878	1 T45884	Human colon specif
2	685	23.0	878	1 V16672	Polynucleotide seq
3	164	3.5	255	1 T22483	Human gene signatu
4	19	0.6	710	1 Q06308	Sequence of DNA fr
5	19	0.6	2184	1 Q06301	Sequence encoding
6	19	0.6	2100	1 V59671	Human secreted pro
7	19	0.6	447	1 V90187	EST clone DH318. N
8	19	0.6	299	1 V89358	EST clone CU493. N
9	18	0.6	3642	1 N50364	DNA fragment encod
10	18	0.6	3474	1 N80908	Sequence of Herpes
11	18	0.6	897	1 Q31988	Guinea pig MBP-1 g
12	18	0.6	3472	1 Q48497	Glycoprotein B (GB
13	18	0.6	735	1 Q73329	Soluble human inte
14	18	0.6	2040	1 T43949	Sequence flanking
15	18	0.6	1386	1 T69545	Rat pheromone rece
16	18	0.6	3473	1 T93651	Herpes simplex vir
17	18	0.6	716	1 V68834	DNA molecule encod
18	18	0.6	14286	1 X13099	Enterococcus faeca
19	17	0.6	18177	1 N90490	DNA of human retin
20	17	0.6	843	1 Q10677	Modified hepatitis
21	17	0.6	2287	1 Q13802	Non-receptor like
22	17	0.6	1560	1 Q29613	Encodes carp Tlata
23	17	0.6	2155	1 Q42418	IDUA - exons I and
24	17	0.6	4475	1 Q45995	IDUA - exons III t
25	17	0.6	13104	1 Q46852	Clone of recombin
26	17	0.6	4529	1 Q46121	Porcine sodium ion
27	17	0.6	314	1 Q60864	Human brain Expres
28	17	0.6	420	1 Q57632	Endothelin-1 (ET-1
29	17	0.6	2129	1 Q63972	Megakaryocyte pote
30	17	0.6	267	1 T20069	Human gene signatu
31	17	0.6	713	1 T31013	Human gene signatu
32	17	0.6	276	1 T26704	Human endothelein-1
33	17	0.6	2404	1 T33220	Oestrogen-regulate
34	17	0.6	80073	1 T58840.5	Continuation (6 of
35	17	0.6	2067	1 T71754	Human alpha-L-idur
36	17	0.6	14507	1 T73568	Expression augment
37	17	0.6	2138	1 T91079	Human CAK1 antigen
38	17	0.6	2454	1 T99071	Partial sequence o
39	17	0.6	680	1 T97392	Aspergillus oryzae
40	17	0.6	1199	1 V42200	CD70 ligand gene u
41	17	0.6	1199	1 V45211	Human endosulfine
42	17	0.6	110000	1 V21209.00	Methanococcus jann
43	17	0.6	110000	1 V21209.02	Continuation (3 of

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB ID	Description
44	17	0.6	110000	1	V21209.03	Continuation (4 of
45	17	0.6	110000	1	V21209.04	Continuation (5 of

Query Match 23.0%; Score 685; DB 1; Length 878;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1992	TGCTCTCAAGGATTTTACAACTTATGACACGATGTAGTATACAGTGTAAAGTGG	2051
1	TGCTCTCAAGGATTTTACAACTTATGACACGATGTAGTATACAGTGTAAAGTGG	60
2052	GGCTCGGGAGAGTTTACAGCAGCAGAGAGATGATACCCAGAGAGTGGACACT	2111
61	GGCTCGGGAGAGTTTACAGCAGCAGAGAGATGATACCCAGAGAGTGGACACT	120
2112	GTACATACCTGGCTGATGAGATGATGAATATACATGGAATCCACAGACCTGAAT	2171
121	GTACATACCTGGCTGATGAGATGATGAATATACATGGAATCCACAGACCTGAAT	180
2172	TATATAGATGATGTTCACACAGCAAGTGTTCACAGCAACATCTCGGAGAGCTC	2231
181	TATATAGATGATGTTCACACAGCAAGTGTTCACAGCAACATCTCGGAGAGCTC	240
2232	ATTGTGGCTTCGATGTCACCAATGCTCCATACCTGATCTTCCAGCTGGCAAT	2291
241	ATTGTGGCTTCGATGTCACCAATGCTCCATACCTGATCTTCCAGCTGGCAAT	300
2292	CACCGACTGAAGGCGGAATTCACGCGGGGAGTCTCATTAATCTGACCTGGACCTCC	2351
301	CACCGACTGAAGGCGGAATTCACGCGGGGAGTCTCATTAATCTGACCTGGACCTCC	360

QY 2352 TGGGATGATTTGACCATGGAACAGCTACAAAGTATATCATTCGATTAAGTACAAGTAT 2411  
DB 361 TGGGGATGATTTATGACCAATGGAACAGCTACAAAGTATATCATTCGATTAAGTACAAGTAT 420  
QY 2412 TCTTGATCTCAGAGACAGTCAATGATATCTCTCAAGTGAATAGTACTAGCTCTCATCC 2471  
DB 421 TCTTGATCTCAGAGACAGTCAATGATATCTCTCAAGTGAATAGTACTAGCTCTCATCC 480  
QY 2472 AAGGAGGCCAACTCTGAGGAAGTCTTTTGTAAACGAGAAAACATTAATCTTTGAAAA 2531  
DB 481 AAGGAGGCCAACTCTGAGGAAGTCTTTTGTAAACGAGAAAACATTAATCTTTGAAAA 540  
QY 2532 TGGCAGAGATCTTTTCATTCGCTATTAGAGCTGTGTAAAGTGCATCTGAATCAGAAAT 2591  
DB 541 TGGCAGAGATCTTTTCATTCGCTATTAGAGCTGTGTAAAGTGCATCTGAATCAGAAAT 600  
QY 2592 ATCCAACTGACAGAGATCTTTTGTATTCCTCCACAGAGCTCCGACAGACACTAG 2651  
DB 601 ATCCAACTGACAGAGATCTTTTGTATTCCTCCACAGAGCTCCGACAGACACTAG 660  
QY 2652 TCCTGATGAACGCTGCTCCTGT 2676  
DB 661 TCCTGATGAACGCTGCTCCTGT 685

RESULT 2  
V16672 2  
ID V16672 standard; cDNA; 878 BP.  
AC V16672;  
DE 22-JUN-1998 (first entry)  
KM Polynucleotide sequence of a colon-specific gene.  
KN Colon-specific gene; probe; detection; expression; human;  
OS Homo sapiens.  
FH Key 2. 685  
FT CDS Location/Qualifiers  
FT /tag= a  
FT /note= "no stop codon given"  
PN US5733748-A.  
PD 31-MAR-1998.  
PF 06-JUN-1995; 469667.  
PR 06-JUN-1995; US-469667.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen C. Yu G.  
DR WPI:98-229823/20.  
PT P-FSDB; W46879.  
PT Colon-specific nucleic acids - useful as probes for detecting colon  
PT Cancer micrometastases  
PS Claim 15; Fig 5A-B; 51pp; English.  
CC V16668-81 represent polynucleotide sequences of partial or full-length  
CC cDNA clones of colon-specific genes. The polynucleotides can be used  
CC as probes to detect expression of the corresponding human genes, e.g. in  
CC diagnostic assays for detecting micrometastases of colon cancer.  
CC Recombinant cells containing the polynucleotides can be used to  
CC produce the polypeptides, in order that antibodies can be raised and  
CC used in further screening or diagnostics.  
CC Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;  
SQ

Query Match 23.0%; Score 685; DB 1; Length 878;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1992 TGTCTACTCAAGGTTTTCACACTTATGACACGATGTGATACAGTGAAGTGGC 2051  
DB 1 TGTCTACTCAAGGTTTTCACACTTATGACACGATGTGATACAGTGAAGTGGC 60  
QY 2052 GGGCTGGAGAGATTAAACGACGACGAGAGAGTGAATACCCGACGAGAGGACACT 2111  
DB 61 GGGCTGGAGAGATTAAACGACGACGAGAGAGTGAATACCCGACGAGAGGACACT 120  
QY 2112 GTACATACCTGGCTGATGAGATGATGAATGATGAATGATGATGATGATGATGATGAT 2171

DB 121 GTACATACCTGGCTGATGAGATGATGAATGATGAATGATGAATGATGATGATGATGAT 180  
QY 2172 TAATAGAGATGATGTTTCAACACAAAGAGTGTGTTTCAGCAGAACATCTCGGAGGCTC 2231  
DB 181 TAATAGAGATGATGTTTCAACACAAAGAGTGTGTTTCAGCAGAACATCTCGGAGGCTC 240  
QY 2232 ATTTGAGTTGATGATGTTCCCAATGCTCCCAATACCTGATCTCTTCCACCTGGCCAAAT 2291  
DB 241 ATTTGAGTTGATGATGTTCCCAATGCTCCCAATACCTGATCTCTTCCACCTGGCCAAAT 300  
QY 2292 CACGACCTGAAGGGGGAATTCACGGGGGAGTCTCATTAATGATGATGATGATGATGATGAT 2351  
DB 301 CACGACCTGAAGGGGGAATTCACGGGGGAGTCTCATTAATGATGATGATGATGATGATGAT 360  
QY 2352 TGGGATGATTAATGACATGGAACAGCTCACAAGATATATCATTTGCAATAGTACAAAT 2411  
DB 361 TGGGATGATTAATGACATGGAACAGCTCACAAGATATATCATTTGCAATAGTACAAAT 420  
QY 2412 TCTTGATCTCAGAGACAGTCAATGATATCTCTCAAGTGAATAGTACTAGCTCTCATCC 2471  
DB 421 TCTTGATCTCAGAGACAGTCAATGATATCTCTCAAGTGAATAGTACTAGCTCTCATCC 480  
QY 2472 AAGGAGGCCAACTCTGAGGAAGTCTTTTGTAAACGAGAAAACATTAATCTTTGAAAA 2531  
DB 481 AAGGAGGCCAACTCTGAGGAAGTCTTTTGTAAACGAGAAAACATTAATCTTTGAAAA 540  
QY 2532 TGGCAGAGATCTTTTCATTCGCTATTAGAGCTGTGTAAAGTGCATCTGAATCAGAAAT 2591  
DB 541 TGGCAGAGATCTTTTCATTCGCTATTAGAGCTGTGTAAAGTGCATCTGAATCAGAAAT 600  
QY 2592 ATCCAACTGACAGAGATCTTTTGTATTCCTCCACAGAGCTCCGACAGACACTAG 2651  
DB 601 ATCCAACTGACAGAGATCTTTTGTATTCCTCCACAGAGCTCCGACAGACACTAG 660  
QY 2652 TCCTGATGAACGCTGCTCCTGT 2676  
DB 661 TCCTGATGAACGCTGCTCCTGT 685

RESULT 3  
T22483 3  
ID T22483 standard; cDNA to mRNA; 255 BP.  
AC T22483;  
DE 22-AUG-1996 (first entry)  
KM Human gene signature H0MS04094.  
KN Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN M09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PI (OKUBA/) OKUBO K.  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 1138-1139; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridize to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.





CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
SQ Sequence 447 BP; 120 A; 107 C; 95 G; 125 T;

Query Match 0.6%; Score 19; DB 1; Length 447;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 529 TACAAATGATGAGAAAT 547  
|||||  
DB 168 TACAAATGATGAGAAAT 186

## RESULT 8

V89358 standard; cDNA; 299 BP.  
AC V89358;  
DT 15-FEB-1999 (first entry)  
DE EST clone C493.  
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
KW gene therapy; ss.  
OS Homo sapiens.  
PN MO9845436-A2.  
PD 15-OCT-1998.  
PF 10-APR-1998; U06955.  
PR 10-APR-1997; US-838821.  
PA (GENE) GENETICS INST INC.  
PI Agostino M.J. Jacobs K. Lavallie ER. McCoy JM. Merberg D.  
PI Racie LA. Spaulding V. Treacy M.  
DR WPI: 99-070077/06.  
PT New Polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.  
PS Claim 1, Page 192; 618bp; English.  
CC The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
SQ Sequence 299 BP; 72 A; 65 C; 48 G; 114 T;

Query Match 0.6%; Score 19; DB 1; Length 299;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 563 GAAGAATACAGCAGTAAG 581  
|||||  
DB 193 GAAGAATACAGCAGTAAG 211

RESULT 9  
N50364/c  
ID N50364 standard; DNA; 3642 BP.  
AC N50364;  
DT 12-NOV-1991 (first entry)  
DE DNA fragment encoding herpes simplex virus 1 glycoprotein B.  
KW Herpes simplex virus 1; glycoprotein B; ss DNA; vaccine.  
OS Herpes simplex virus 1.  
FH Key Location/Qualifiers  
FT cds 438..3149  
PN W08504587-A.

PD 24-OCT-1985.  
PF 04-APR-1985; U00587.  
PR 06-APR-1984; US-587784.  
PR 17-JUL-1984; US-631659.  
PI (CHIR-) Chiron Corp.  
PI Burke RL, Pachl C, Valenzuela PDT, Urdea MS;  
DR WPI: 85-276087/44.  
DR P-PSDB; P50312.

PT Recombinant herpes simplex vaccine - prep. by expression of DNA  
PT constructs in a eukaryotic host.  
PS Disclosure; Table 1 page 26-30; 80pp; English.  
CC Herpes simplex virus glycoprotein B or fragments may be used in a  
CC vaccine against HSV. Dosage is 10 micrograms to 2 mg/kg. The  
CC glycoprotein DNA is expressed in an eukaryotic host, esp.  
CC Saccharomyces cerevisiae, CHO cells and COS cells. Suitable plasmids  
CC are pYHS115, 116, 117, 118 and 119.  
SQ Sequence 3642 BP; 689 A; 1241 C; 1105 G; 607 T;

Query Match 0.6%; Score 18; DB 1; Length 3642;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 641 ACACCAAAAGATGCACAT 658  
|||||  
DB 3209 ACACCAAAAGATGCACAT 3192

## RESULT 10

N80908/c  
ID N80908 standard; DNA; 3474 BP.  
AC N80908;  
DT 09-MAR-1992 (first entry)  
DE Sequence of Herpes simplex virus (HSV) glycoprotein B (gB) 2 DNA.  
OS Vaccine; herpes simplex virus; therapy; ss.  
OS Herpes simplex virus.  
FH Key Location/Qualifiers  
FT cds 309..3023  
PN W08802634-A.  
PD 21-APR-1988.  
PF 20-OCT-1987; U02709.  
PR 20-OCT-1986; US-921213.  
PR 20-JUL-1987; US-079605.  
PA (CHIR-) CHIRON CORP.  
PI Burke LR, Pachl C, Valenzuela P;  
DR WPI: 88-119368/17.  
DR P-PSDB; P80915.  
PT Vaccine for treatment of herpes simplex virus - contains  
PT recombinant HSV glyco:proteins B and D  
PS Example; Fig 4; 71pp; English.  
CC Prepn. of recombinant gB and gD is described in W088504587. The  
CC amino acid sequence and DNA sequence for gB 1 presented in P80914  
CC and N80907 differ from that originally presented in Table 1 of  
CC International Publication No. WO 85/04587. The DNA sequence in Table  
CC 1 contains an error in that an additional nucleotide (G) is listed  
CC at position 607 which resulted in a shift in reading frame relative  
CC to N80907 from which this nucleotide has been deleted.  
SQ Sequence 3474 BP; 666 A; 1183 C; 1058 G; 567 T;

Query Match 0.6%; Score 18; DB 1; Length 3474;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 641 ACACCAAAAGATGCACAT 658  
|||||  
DB 3086 ACACCAAAAGATGCACAT 3069

RESULT 11  
Q31988/c  
ID Q31988 standard; cDNA; 897 BP.

AC 031988: (first entry)  
DI 30-APR-1993  
DE Guinea pig MBP-1 gene.  
KW Major basic protein; anti-asthma; ss.  
OS Cavia porcellus.  
FH Key  
FT cds Location/Qualifiers  
FT /tag= a  
FT signal\_peptide 75..119  
FT /tag= b  
FT signal\_peptide 120..416  
FT /tag= c  
FT /note= "precursor peptide"  
FT mat\_peptide 417..773  
FT /tag= d  
PN J04320686-A.  
PD 11-NOV-1992.  
PF 22-APR-1991: 090704.  
PR 22-APR-1991: JP-090704.  
PA (Saka) OTSUKA PHARM CO LTD.  
DR WPI: 92-426680/52.  
P-PSDB: R29716.  
PT A guinea pig MBP gene - contains DNA base sequence coding MBP  
PS gene having 1 of 2 specified aminoacid sequences  
PS Claim 1: page 15: 20pp; Japanese.  
CC The sequence shows the cDNA sequence of the guinea pig major basic  
CC protein (MBP)-1 gene. The sequence was obt. by first isolating MBP-1  
CC and MBP-2 from the abdominal transudate of male guinea pigs treated  
CC with polymyxin and ascaris extract. MBP-1 and MBP-2 were used to  
CC design probes for screening of a guinea pig eosinophil cDNA library.  
CC The MBPs may be pred. in large amounts by this method and may be  
CC used for the confirmation of the effect of an anti-asthma drug.  
CC See also Q31989-93.  
SQ Sequence 897 BP: 201 A: 232 C: 282 G: 182 T:

Query Match 0.6%; Score 18; DB 1; Length 897;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 227 CAGGAAGCGATTATT 244  
DB 876 CAGGAAGCGATTATT 859  
RESULT 12  
Q48497/c  
ID Q48497 standard; DNA: 3472 BP.  
AC Q48497;  
DT 29-MAR-1994 (first entry)  
DE Glycoprotein B (gB2)  
KW Glycoprotein; gB1; gB2; herpes simplex virus; HSV-1; HSV-2;  
KW flanking: initiation; termination; transcription; translation;  
KW regulatory sequence; ss.  
OS Herpes simplex virus 2 strain 333.  
FH Key  
FT cds Location/Qualifiers  
FT /tag= a  
FT /tag= a  
PN US5244792-A.  
PD 14-SEP-1993.  
PF 06-APR-1984: 597784.  
PR 06-APR-1984: US-597784.  
PR 20-OCT-1986: US-921730.  
PR 20-SEP-1990: US-587179.  
PA (CHIR) CHIRON CORP.  
PI Burke RL, Pacht C, Valenzuela PDT;  
DR WPI: 93-302641/38.  
P-PSDB: R41779.  
PT DNA construct for expressing HSV glycosylated polypeptide -  
PT useful for vaccinating against HSV-1 and -2 infections  
PS Disclosure: Fig 4; 33pp; English.  
CC DNA constructs for expressing a glycosylated polypeptide in a  
CC eukaryotic host cell comprises (1) an oligonucleotide sequence (os)

CC free of natural flanking sequences, encoding glycoprotein B (gB)  
CC of HSV, or its fragments, and (2) 5' initiating and 3' terminating  
CC transcriptional and translational regulatory sequences flanking OS,  
CC at least one of these sequences not being from HSV. The HSV-1 and  
CC HSV-2 sequences are given in Q48496 and Q48497 respectively.  
SQ Sequence 3472 BP: 661 A: 1193 C: 1053 G: 565 T;

Query Match 0.6%; Score 18; DB 1; Length 3472;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 ACACCAAAAGATGCACAT 658  
DB 3083 ACACCAAAAGATGCACAT 3066

RESULT 13  
Q73229/c  
ID Q73229 standard; cDNA: 735 BP.  
AC Q73229;  
DT 11-APR-1995 (first entry)  
DE Soluble human interferon gamma receptor coding sequence.  
KW Interferon; gamma; IFN; receptor; immunoglobulin; constant domain;  
KW light chain; heavy chain; Ig; chimeric protein; fusion protein;  
KW autoimmune disease; chronic inflammation; allotransplant; rejection;  
KW multiple sclerosis; fulminant hepatitis; neurological disease; AIDS;  
KW poliovirus; Lyme disease; septicemia; treatment; therapy;  
KW delayed type hypersensitivity; ss.  
OS Homo sapiens.  
FH Key  
FT cds Location/Qualifiers  
FT /tag= a  
FT /product= Soluble interferon gamma receptor.  
FT signal\_peptide 1..51  
FT /tag= b  
FT mat\_peptide 52..735  
FT /tag= c  
PN EP-614981-A.  
PD 14-SEP-1994.  
PF 18-FEB-1994: 102452.  
PR 05-MAR-1993: EP-810170.  
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.  
PI Dembic Z, Garotta G, Gentz R;  
DR WPI: 94-281208/35.  
P-PSDB: R62023.  
PT Chimeric human interferon-gamma receptor/immunoglobulin proteins  
PT - used to inhibit binding of interferon-gamma to its specific  
PT receptor in the treatment of illnesses  
PS Disclosure: Figure 1: 29pp; English.  
CC The soluble form of the Interferon (IFN) gamma receptor comprises  
CC the whole extracellular domain of the natural receptor from the N-  
CC terminus to the transmembrane region, lacks the cytoplasmic and  
CC transmembrane domains of the natural receptor and specifically binds  
CC IFN-gamma. The sequence encoding the soluble IFN-gamma receptor can  
CC be used in constructs encoding chimeric proteins where the other  
CC component of the chimeric protein is part of all of the constant  
CC domain of a human immunoglobulin heavy or light chain. The  
CC recombinant proteins can be used to inhibit IFN-gamma binding to its  
CC specific receptor. They can be used for the treatment of  
CC illnesses, especially autoimmune diseases, chronic inflammation,  
CC delayed type hypersensitivity, allotransplant rejections, multiple  
CC sclerosis, fulminant hepatitis, inflammatory neurological diseases  
CC and neurological complications of AIDS, poliovirus infections, Lyme  
CC disease and septicemia. The presence of the immunoglobulin  
CC component in the chimeric protein increases the proteins half life in  
CC vivo.  
SQ Sequence 735 BP: 223 A: 135 C: 170 G: 207 T:

Query Match 0.6%; Score 18; DB 1; Length 735;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 529 TACATAATGATGAGAAA 546  
 DB 256 TACATAATGATGAGAAA 239

## RESULT 14

T43949  
 ID T43949 standard; DNA: 2040 BP.  
 AC T43949;  
 DT 18-AUG-1997 (first entry)  
 DE Sequence flanking marker 63-2 in HH region of chromosome 6p2.1.  
 KM Primer: polymerase chain reaction; amplify: hereditary haemochromatosis;  
 KM HH: mutation, HH-associated allele; base-pair polymorphism; HHP-1;  
 KM HHP-19; HHP-29; microsatellite repeat allele; genetic marker;  
 KM interferon treatment; hepatitis C infection; ss.  
 OS Synthetic.  
 PN MO9635803-A1.  
 PD 14-NOV-1996.  
 PF 08-MAY-1996; U06583.  
 PR 08-MAY-1995; US-436074.  
 PR 15-NOV-1995; US-553032.  
 PR 09-FEB-1996; US-599252.  
 PA (MERC-) MERCATOR GENETICS INC.  
 PI Drayna DF, Feder JN, Gullike A, Kimmel BE, Thomas WJ;  
 PI Wolff RK.  
 DR WPI: 96-518691/51.  
 PT Diagnosing and genotyping of hereditary haemochromatosis (HH) -  
 PT using primers to detect specific polymorphisms of the HH gene on  
 PT chromosome 6p2.1 or novel microsatellite markers  
 PS Disclosure: Fig 1Y: 67pp; English.  
 CC The sequences given in T43925-55 represent portions of the genome  
 CC surrounding several markers of the invention. The markers were  
 CC identified using the series of primer pairs given in T71901-72  
 CC which were used to determine the presence or absence of the common  
 CC hereditary haemochromatosis (HH) gene mutation in an individual. The  
 CC method comprised assessing genomic DNA from an individual for the  
 CC presence or absence of the HH-associated allele of the single base-pair  
 CC polymorphism HHP-1, HHP-19 or HHP-29, and/or at least one non-optional  
 CC marker comprising the following microsatellite repeat alleles of group  
 CC A and optionally of group B:  
 CC Group A: 1909, 1884, 1A2, 1E4, 24E2, 2B8, 3321-1, 4073-1, 4440-1, 4440-2,  
 CC 711-1, 5091-1, 3216-1, 4072-2, 950-1, 950-3, 950-4, 950-5, 950-6,  
 CC 950-8, 63-2, 63-3, 65-1, 65-2, 373-8, 373-29, 68-1, 241-6, 241-29;  
 CC Group B: D6S464, D6S306, D6S258, D6S265, D6S105 and D6S1001.  
 CC The absence of the genotype indicates the likelihood of the presence of  
 CC the HH mutation. Knowledge of the new genetic markers allows the  
 CC definition of genotypes characteristic of heterozygous carriers and  
 CC homozygotes having a HH mutation in their genomic DNA. The potential for  
 CC HH in an individual interferes with the effectiveness of interferon  
 CC treatment for hepatitis C infection. By diagnosing this potential, the  
 CC responsiveness of interferon treatment may be evaluated.  
 SO Sequence 2040 BP: 713 A: 409 C: 398 G: 513 T:

Query Match 0.6%; Score 18; DB 1; Length 2040;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1647 CTTGTAGTGACAAA 1664  
 DB 195 CTTGTAGTGACAAA 212

## RESULT 15

T69545  
 ID T69545 standard; DNA: 1386 BP.  
 AC T69545;  
 DT 26-AUG-1997 (first entry)  
 DE Rat pheromone receptor VNI cDNA.  
 KM Pheromone receptor; vomeronasal sensory neuron; social behaviour;  
 KM maternal behaviour; reproductive behaviour; fertility;  
 KM hormone secretion; ss.

OS Rattus sp.  
 FH Key Location/Qualifiers  
 FT cds 270..1217  
 FT /\*tag= a

PN MO9714790-A1.  
 PD 24-APR-1997.  
 PF 18-OCT-1996; U16637.  
 PR 19-OCT-1995; US-005698.  
 PA (VXCO ) UNIV COLOMBIA NEW YORK.  
 PI Axel R, Dulac C;  
 PI WPI: 97-245107/22.  
 DR P-PSDB: W19103.  
 PT Nucleic acid molecule encoding vertebrate pheromone receptor -  
 PT useful to identify modulators for control of reproductive and social  
 PT behaviour, fertility and hormone secretion  
 PS Claim 1; Fig 8; 123pp; English.  
 CC A cDNA clone (T69545) codes for rat pheromone receptor VNI  
 CC (W19103), a member of a novel family of presumed 7-transmembrane  
 CC domain receptors that are evolutionary independent of the odorant  
 CC receptors of the main olfactory epithelium (MOE). It was isolated  
 CC by differential screening of cDNA libraries constructed from rat  
 CC single vomeronasal sensory neurons. A VNI insert in pBluescript  
 CC has been deposited as ATCC 97294. VNI is expressed only in about  
 CC 4% of VSN and never in the MOE. cDNA clones for VN2-7 were obt.  
 CC by PCR and hybridisation (see also T69546-50. VN2 sequence not  
 CC given). VN polypeptides (W19103-09) can be expressed in host-  
 CC vector systems for use in identifying modulators for control of  
 CC maternal, reproductive and social behavior, to increase fertility,  
 CC control hormone secretion and to regulate food uptake in humans and  
 CC animals. A human homologue, H625 (T69551), has also been obt.  
 SO Sequence 1386 BP: 357 A: 322 C: 271 G: 435 T:

Query Match 0.6%; Score 18; DB 1; Length 1386;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1342 TCTGCAGCTCAAGACCTA 1359  
 DB 1042 TCTGCAGCTCAAGACCTA 1059

Search completed: August 5, 1999, 17:42:22  
 Job time: 6377 sec

This Page Blank (uspto)

PT New isolated nucleic acids from *Streptococcus pneumoniae* - useful  
PT e.g. for identifying anti-bacterial(s) for treatment and prevention  
PT of meningitis  
PS Claim 71, 130pp; English.  
CC The sequence is that of a coding region isolated from  
CC *S. pneumoniae*. Its encoded protein, or agonists of it,  
CC may be useful as an antibacterial for treatment or  
CC prevention of infection, specifically caused by *S. pneumoniae*  
CC (particularly meningitis), but possibly also *Helicobacter*  
CC pylori (ulcers and gastric cancer). It may be of particular  
CC use before insertion of an in-dwelling device or any other  
CC invasive procedure. The protein, or nucleic acid encoding  
CC it, can also be used in vaccines to induce a cellular  
CC and/or humoral immune response or to screen for other

88 AATGATGAATACAAATGGAATCCACCAAGACCTGAAATTAAATAAGGATGATGTTCAACAC 147

```
Db 6280 ATTCTTGAATTCATTACTTACCAATTCCTCCTATTAGAGAGAAAGATGACAT 6339
Oy 148 AAGCAAGTGTGT 159
Db 6340 AAGAAGCGCTGT 6351

RESULT 8
T30194
ID T30194 standard; cDNA: 6371 BP.
AC T30194:
DE 25-OCT-1996 (first entry)
DE Peripheral nervous system sodium channel peptide-1 alpha-subunit gene.
KW Human; peripheral nervous system; sodium channel; PNIA;
KW dorsal root ganglion; sodium-agonist; sodium-antagonist;
KW drug screening; analgesic; hypotensive; antiinflammatory; trauma;
KW pain; neurological disorder; antisense; gene therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 49..5982
FT /tag= a
FT /product= PNIA protein
PN M09614077-A1.
PD 17-MAY-1996.
PF 02-NOV-1995; U14251.
PR 02-NOV-1994; US-334029.
PR 07-JUN-1995; US-482401.
PA (TROP-) TROPHIX PHARM INC.
PA (UNYV) UNIV NEW YORK STATE RES FOUND.
PI Borden LA, Halegoua S, Mandel G;
DR WPI: 96-251547/25.
DR P-PSDB: R99640.
PT Nucleic acid encoding peripheral nervous system specific sodium
PT channel peptide - useful for sodium channel-associated disease or
PT trauma.
PS Example 4; Fig 13; 80pp; English.
CC The sequence encodes a human peripheral nervous system sodium channel
CC peptide-1 alpha-subunit (PNIA), with sodium channel activity. A
CC probe has been isolated by PCR from a human peripheral nervous
CC system cDNA library, using probes derived from the rat PNI protein
CC sequence (R99638-39), and has been used to screen a human dorsal
CC root ganglion cDNA library, to give the full-length sequence, using
CC PCR primers T30198-99 and T18167. A related sequence (T30195) has
CC also been isolated. The sodium channel peptide may be used to
CC isolate sodium-agonists and sodium-antagonists for use as
CC analgesics, hypotensives, antiinflammatorys, and in therapy of
CC sodium channel-associated pathology or trauma, e.g. neurological
CC disorders. The DNA (in sense or antisense orientation) may be used
CC in gene therapy.
SQ Sequence 6371 BP; 1948 A; 1231 C; 1374 G; 1818 T;

Query Match 12.4%; Score 28.2; DB 1; Length 6371;
Best Local Similarity 53.1%; Pred. No. 8.6;
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Oy 45 ACCCCAGAGAGTGGAGCACTGTACATACCTGCTGGATTGGAATGATGAATACATG 104
Db 3382 AGCTCCTCAGAGTGGAGCACTGTATACCTTGGCTGGAGAGAGAAAGACAG 3441
Oy 105 GAATCCACCAGACCTGAATTAATAGATGATGTTCAACCAAGCAAGTGT 157
Db 3442 GCTGAACCTATGAATTCGATGAGCCAGAGGCTGTTTCACAGATGTTGTGT 3494

RESULT 9
T30195
ID T30195 standard; cDNA: 6404 BP.
AC T30195:
DE 25-OCT-1996 (first entry)
DE Peripheral nervous system sodium channel peptide-1 alpha-subunit gene.
KW Human; peripheral nervous system; sodium channel; PNIB;
```

```
KW dorsal root ganglion; sodium-agonist; sodium-antagonist;
KW drug screening; analgesic; hypotensive; antiinflammatory; trauma;
KW pain; neurological disorder; antisense; gene therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 49..6015
FT /tag= a
FT /product= PNIB protein
FT misc_difference 334..336
FT /tag= b
FT /codon-seq=ACA, aa:Ala
FT misc_difference 1738..1740
FT /tag= c
FT /codon-seq=ATA, aa:Leu
PN M09614077-A1.
PD 17-MAY-1996.
PF 02-NOV-1995; U14251.
PR 02-NOV-1994; US-334029.
PR 07-JUN-1995; US-482401.
PA (TROP-) TROPHIX PHARM INC.
PA (UNYV) UNIV NEW YORK STATE RES FOUND.
PI Borden LA, Halegoua S, Mandel G;
DR WPI: 96-251547/25.
DR P-PSDB: R99641.
PT Nucleic acid encoding peripheral nervous system specific sodium
PT channel peptide - useful for sodium channel-associated disease or
PT trauma.
PS Example 4; Fig 14; 80pp; English.
CC The sequence encodes a human peripheral nervous system sodium channel
CC peptide-1 alpha-subunit (PNIB), with sodium channel activity. A
CC probe has been isolated by PCR from a human peripheral nervous
CC system cDNA library, using probes derived from the rat PNI protein
CC sequence (R99638-39), and has been used to screen a human dorsal
CC root ganglion cDNA library, to give the full-length sequence, using
CC PCR primers T30198-99 and T18167. A related sequence (T30194) has
CC also been isolated. The sodium channel peptide may be used to
CC isolate sodium-agonists and sodium-antagonists for use as
CC analgesics, hypotensives, antiinflammatorys, and in therapy of
CC sodium channel-associated pathology or trauma, e.g. neurological
CC disorders. The DNA (in sense or antisense orientation) may be used
CC in gene therapy.
SQ Sequence 6404 BP; 1961 A; 1236 C; 1381 G; 1826 T;

Query Match 12.4%; Score 28.2; DB 1; Length 6404;
Best Local Similarity 53.1%; Pred. No. 8.6;
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Oy 45 ACCCCAGAGAGTGGAGCACTGTACATACCTGCTGGATTGGAATGATGAATACATG 104
Db 3415 AGCTCCTCAGAGTGGAGCACTGTATACCTTGGCTGGAGAGAGAAAGACAG 3474
Oy 105 GAATCCACCAGACCTGAATTAATAGATGATGTTCAACCAAGCAAGTGT 157
Db 3475 GCTGAACCTATGAATTCGATGAGCCAGAGGCTGTTTCACAGATGTTGTGT 3527

RESULT 10
V27333
ID V27333 standard; DNA: 868 BP.
AC V27333:
DE 02-OCT-1998 (first entry)
DE Streptococcus pneumoniae SP0013 nucleotide.
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis; ss.
OS Streptococcus pneumoniae.
FH Key Location/Qualifiers
FT cds 2..868
FT /tag= a
FT /product= "SP0013"
FT /note= "no stop codon given"
PN M09818930-A2.
PD 07-MAY-1998.
```





Db 110 CATCTCGGTAGATACAGT 92

RESULT 15

Continuation (5 of 6) of T58840 from base 400001 (Mycoplasma genitalium genome.)  
T58840\_4/c  
WP Sequence split into 6 fragments LOCUS T58840 Accession T58840

WP	Fragment Name	Begin	End
WP	T58840_0	1	110000
WP	T58840_1	100001	210000
WP	T58840_2	200001	310000
WP	T58840_3	300001	410000
WP	T58840_4	400001	510000
WP	T58840_5	500001	580073

Query Match 12.2% Score 27.8; DB 1; Length 110000;  
Best Local Similarity 62.0%; Pred. No. 39;  
Matches 44; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 83 TTGAGATGATGAATACATGCAATCCACCAAGACCTGAATTAAATAGATGATGTC 142  
DB 86536 TTGAAATATATGAGTCAACCTAAAGAAACAGAGCTGAATTACTTTGATGAACCTA 86477  
QY 143 AACCAACCAA 153  
DB 86476 AAGAACTTCAA 86466

Search completed: August 6, 1999, 04:03:46  
Job time: 1578 sec

**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 1998 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:52 ; Search time 3059.95 Seconds  
(without alignments)  
146.331 Million cell updates/sec

Title: US-09-049-696-13

Perfect score: 227  
Sequence: 1 GTGGCGGCTGTGGAGAGT.....AATGCTCCATACCTGATCT 227

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	59	26.0	734	49	A1660957	A1660957 wf20d08.x
2	19	8.4	516	30	AA255386	AA255386 SW13CA242
3	19	8.4	321	40	AA933120	AA933120 SWEM138A0
c 4	18	7.9	535	37	AA668060	AA668060 SMOV3KCCAM
c 5	17	7.5	416	20	D27036	D27036 CEIK009C7F
c 6	17	7.5	360	20	D34790	D34790 ATTS5834 Gr
7	17	7.5	213	24	F19795	F19795 ATTS5834 Gr
8	17	7.5	270	26	W82028	W82028 mF02e01.r1
9	17	7.5	292	27	AA049828	AA049828 mJ09a09.r
10	17	7.5	361	27	AA050569	AA050569 mJ16b12.r
11	17	7.5	310	28	AA123834	AA123834 mp93c04.r
12	17	7.5	331	30	AA220617	AA220617 my25f11.r
c 13	17	7.5	356	30	AA250523	AA250523 mz60f03.r
c 14	17	7.5	506	30	AA266134	AA266134 mz49c10.r
c 15	17	7.5	360	36	C63765	C63765 C63765 Yuj1
c 16	17	7.5	376	36	C64568	C64568 C64568 Yuj1
c 17	17	7.5	460	37	AA718536	AA718536 vE84h01.r
c 18	17	7.5	552	38	AA772272	AA772272 a142c09.s
c 19	17	7.5	654	39	AA842640	AA842640 MBAPC28G0
c 20	17	7.5	316	39	AA881199	AA881199 vx10a07.r
21	17	7.5	293	40	AA915649	AA915649 vz33a03.r
22	17	7.5	513	40	AA959963	AA959963 vw53b08.s
23	17	7.5	433	40	AA986673	AA986673 uc82q03.y
24	17	7.5	340	42	AA155561	AA155561 ue05e12.r
25	17	7.5	456	44	AI286582	AI286582 UK04c11.y
26	17	7.5	396	47	AI662101	AI662101 ub70a08.x
27	17	7.5	554	49	AI642202	AI642202 vw53b08.x
28	17	7.5	414	51	AT716936	AT716936 UT-R-Y0-a
c 29	16	7.0	339	20	D32510	D32510 CEIK018A2R
c 30	16	7.0	261	20	M78092	M78092 EST01682 Su
c 31	16	7.0	336	24	D65670	D65670 CEIK065H2R
c 32	16	7.0	360	24	D68012	D68012 CEIK022E7F
c 33	16	7.0	360	24	D74434	D74434 CEIK081A2F
c 34	16	7.0	556	24	H91960	H91960 y862b02.r1
c 35	16	7.0	255	25	N84623	N84623 J0301F HumA
c 36	16	7.0	240	27	C09842	C09842 C09842 Yuj1
c 37	16	7.0	300	27	C10238	C10238 C10238 Yuj1
c 38	16	7.0	462	28	AA074068	AA074068 zm75b05.s
c 39	16	7.0	704	29	AA141002	AA141002 CK01076.5
c 40	16	7.0	393	29	AA160427	AA160427 zo74c03.r
c 41	16	7.0	320	31	AA284330	AA284330 zs36g03.s
c 42	16	7.0	462	31	AA303727	AA303727 EST16395
43	16	7.0	394	33	AA404938	AA404938 28332 Lam
44	16	7.0	479	35	AA541829	AA541829 vJ01f01.r
c 45	16	7.0	420	35	AA588646	AA588646 nj30f11.s

#### ALIGNMENTS

RESULT 1  
A1660957/c A1660957 734 bp mRNA EST 10-MAY-1999  
LOCUS wf20d08.x1 Soares.Dieckgraefe.colon\_NHUC Homo sapiens CDNA clone  
DEFINITION IMAGE:2351151 3' similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA  
sequence.  
ACCESSION A1660957  
NID 94764340

VERSION  
 EST.  
 AI660957.1 GI:4764540  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 human  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE  
 AUTHORS  
 TITLE  
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 Unpublished (1997)  
 On Mar 20, 1998 this sequence version replaced gi:2980033.

JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium ([info@image.jnl.gov](mailto:info@image.jnl.gov)) for further information.  
 Seq primer: -400p from Glbco  
 High quality sequence stop: 477.  
 Location/Qualifiers  
 1..734  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="x"  
 /clone="IMAGE:2351151"  
 /clone\_11b="Soares\_Dieckgraefe\_colon\_NHUC"  
 /tissue\_type="colonic mucosa from 5 ulcerative colitis patients"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="organ: colon; Vector: pRT19-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTCACATCTGAGAGTGACGGCCCTCCTTCTTTTCTTTT 3'), double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@m.wustl.edu); colonic mucosa represents a range of disease involvement from mild cryptitis to severe ulceration, fibrosis, and degeneration. Library constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT  
 200 a 143 c 159 g 228 t 4 others

ORIGIN

Query Match 26.0%; Score 59; DB 49; Length 734;  
 Best Local Similarity 100.0%; Pred. No. 2,4e-23;  
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

Cy 169 ACATCTCGGGAGGCGCATTTGGCTTGATGATGCCAAATGCTCCCTACTGATCT 227  
 |||||||  
 Db 602 ACATCTCGGGAGGCGCATTTGGCTTGATGATGCCAAATGCTCCCTACTGATCT 544

RESULT 2  
 LOCUS AA255386 516 bp mRNA EST 14-MAR-1997  
 DEFINITION SW31CA2426SK Brugia malayi infective larva cDNA (SW94W-L-BM3)  
 ACCESSION AA255386 Brugia malayi cDNA clone SW31CA2426 5', mRNA sequence.  
 NID g1889989  
 VERSION AA255386.1 GI:1889989  
 KEYWORDS EST.  
 SOURCE Brugia malayi.  
 ORGANISM Brugia malayi.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida; Filarioidea; Onchocercidae; Brugia.  
 REFERENCE 1 (bases 1 to 516)  
 AUTHORS Williams,S.A.  
 TITLE Genes expressed in infective third stage larvae of Brugia malayi  
 JOURNAL Unpublished,(1995)

On Dec 30, 1996 this sequence version replaced gj:1529606.

COMMENT

Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu

Seq primer: pBluescript SK.  
Location/Qualifiers

FEATURES

Source

1. 516  
/organism="Brugia malayi"  
/strain="TR3 Labs"  
/db\_xref="taxon:6279"  
/clone="SM31CA2426"  
/clone\_lib="Brugia malayi infective larva cDNA  
(SA94W/-Bml3)"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: lambda Unizap XR; Site\_1: EcoR I; Site\_2:  
Xho I; Lymphatic filarial nematode parasite of humans.  
mRNA was prepared from third stage infective larvae of  
Brugia malayi isolated from mosquitoes 10 days after  
infection and converted to double stranded cDNA using  
reverse transcriptase and oligo(dt) followed by Kase H  
and DNAPol I. The library had 1.6 x 10E6 independent  
recombinants and average insert size was 900 base pairs.  
The library was constructed by Wenhong Lu. The library is  
available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 199 a 97 c 74 g 141 t 5 others

ORIGIN

Query Match 8.4%; Score 19; DB 30; Length 516;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GAAATTATAAGGATGATG 139  
|||||  
Db 226 GAAATTATAAGGATGATG 244

RESULT 3  
AA933120 321 bp mRNA EST 24-APR-1998  
LOCUS SM5ML35A05733 Brugia malayi U3 subcloned cDNA library  
DEFINITION (SAM97LS-Bml3SA) Brugia malayi cDNA clone SM5ML35A057 5' or 3', mRNA  
sequence.  
ACCESSION AA933120  
NID 93087474  
VERSION AA933120.1 GI:3087474  
KEYWORDS EST.  
SOURCE Brugia malayi.  
ORGANISM Brugia malayi  
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;  
Filarioidea; Onchocercidae; Brugia.  
1 (bases 1 to 321)  
Saunders, L. and Williams, S.A.  
Genes expressed in infective third stage larvae of Brugia malayi  
(1998)  
Unpublished (1998)  
On Sep 29, 1997 this sequence version replaced gj:1520597.

JOURNAL COMMENT

Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu  
Seq primer: T3.

FEATURES  
source

Location/Qualifiers  
1. 321  
/organism="Brugia malayi"  
/db\_xref="taxon:6279"  
/clone="SMBML3A057"  
/clone\_lib="Brugia malayi L3 subtracted cDNA library (SAM97LS-Bml3SA)"  
/dev\_stage="L3 cDNA subtracted with microfilaria, adult male and adult female cDNA libraries"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: Lambda ZAP Express; Site 1: Eco RI; Lymphatic filarial nematode parasite of humans. cDNA was amplified by PCR using T3 and T7 primers from conventionally constructed libraries: L3 (SAM94WL-Bml3), microfilaria (SAM94LS-Bmlf), adult male (SAM94NL-Bmlm) and adult female (SAM96MLM-Bmaf). Microfilaria, adult male and adult female cDNAs were subtracted from L3 cDNA using the PCR-select cDNA subtraction kit (Clontech). The library has 2.6 x 10<sup>6</sup> independent recombinants and the average insert size is ~500bp. The library was constructed by Lori Saunders. The library is available from Dr. S. A. Williams, email: genome@smith.edu"

## BASE COUNT

116 a 69 c 48 g 88 t

## ORIGIN

Query Match 8.4%; Score 19; DB 40; Length 321;

Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GAAATTATTAAGATGATG 139

|||||

DB 122 GAAATTATTAAGATGATG 140

## RESULT 4

## LOCUS

AA668060/c 535 bp mRNA EST 20-NOV-1997

## DEFINITION

SMOV3MCA02B01SK Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCA02B01 5', mRNA sequence.

## ACCESSION

## NID

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Onchocerca volvulus.  
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.  
1 (bases 1 to 535)

REFERENCE  
AUTHORS Williams, S.A., Lizotte-Waniewski, M., laney, S. and Lustigman, S.  
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus  
JOURNAL Unpublished (1997)  
COMMENT On May 9, 1995 this sequence version replaced gl:002343.

## JOURNAL

## COMMENT

FEATURES  
source

Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu  
Seq primer: Bluescript SK.  
Location/Qualifiers  
1. 535  
/organism="Onchocerca volvulus"  
/strain="Kumba, Cameroon"  
/db\_xref="taxon:6282"  
/map="1"  
/clone="SMOV3MCA02B01"  
/clone\_lib="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovml3)"

/dev\_stage="molting L3"  
/lab\_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10<sup>6</sup> independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@bc.org)."

BASE COUNT 178 a 133 c 109 g 112 t 3 others  
ORIGIN

Query Match 7.9%; Score 18; DB 37; Length 535;

Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 CCAATGCTCCCATCT 222

|||||

DB 286 CCAATGCTCCCATCT 269

## RESULT 5

## LOCUS

D27036 416 bp mRNA EST 20-NOV-1995

## DEFINITION

CEAK009C7F Yuiji Kohara unpublished cDNA Caenorhabditis elegans CDNA clone YK9C7 5', mRNA sequence.

## ACCESSION

## NID

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditina; Rhabditidae; Pelodierinae; Caenorhabditis.  
1 (bases 1 to 416)

REFERENCE  
AUTHORS Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.  
TITLE Toward an expression map of the C. elegans genome  
JOURNAL Unpublished (1994)  
COMMENT

## JOURNAL

## COMMENT

Contact: Yuiji Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 0559-75-0771  
Fax: 0559-75-6240  
Email: ykohara@dbj.nig.ac.jp.

FEATURES  
source

Location/Qualifiers  
1. 416  
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(e1489)"  
/db\_xref="taxon:6239"  
/clone="YK9C7"  
/clone\_lib="Yuiji Kohara unpublished cDNA"  
/note="dev\_stage=varied, sex=Hermaphrodite male, tissue\_type=whole animal."  
BASE COUNT 148 a 65 c 99 g 103 t 1 others  
ORIGIN

Query Match 7.5%; Score 17; DB 20; Length 416;  
Best Local Similarity 100.0%; Pred. No. 19;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 ATGATGTTCAACACAG 150  
 |||||||  
 Db 245 ATGATGTTCAACACAG 229

RESULT 6  
 D34790/c 360 bp mRNA EST 08-AUG-1994  
 LOCUS CELK012F1F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA  
 DEFINITION clone yk12f1 5', mRNA sequence.  
 ACCESSION D34790  
 NID D34790  
 KEYWORDS 6526298  
 VERSION D34790.1 GI:526298  
 SOURCE EST.  
 ORGANISM Caenorhabditis elegans.  
 Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 360)  
 Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and  
 Tabara,H.  
 Toward an expression map of the C.elegans genome  
 Unpublished (1994)

TITLE Contact: Yuji Kohara  
 JOURNAL Gene Library Lab  
 COMMENT National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel: 0559-75-0771  
 Fax: 0559-75-6240  
 Email: ykohara@dbj.nig.ac.jp.

FEATURES  
 Source location/Qualifiers  
 1..360  
 /organism="Caenorhabditis elegans"  
 /strain="CB1489 him-8(e1489)"  
 /db\_xref="taxon:6239"  
 /clone="yk12f1"  
 /clone\_lib="Yuji Kohara unpublished cDNA"  
 /note="dev\_stage=varied; sex=Hermaphrodite male;  
 tissue\_type=whole animal"

BASE COUNT 128 a 57 c 81 g 90 t 4 others

ORIGIN

Query Match 7.5%; Score 17; DB 20; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 ATGATGTTCAACACAG 150  
 |||||||  
 Db 194 ATGATGTTCAACACAG 178

RESULT 7  
 F19795 213 bp mRNA EST 16-JAN-1996  
 LOCUS ATTS5834 Grenoble-B Arabidopsis thaliana cDNA clone GBGF814 3',  
 DEFINITION mRNA sequence.  
 ACCESSION F19795  
 NID g1154996  
 VERSION F19795.1 GI:1154996  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Eurosidae II; Brassicales; Brassicaceae;  
 Arabidopsis.  
 1 (bases 1 to 213)  
 CNRS.  
 The Arabidopsis thaliana transcribed genome: the GDR cDNA program

REFERENCE  
 AUTHORS  
 TITLE

JOURNAL Unpublished (1996)  
 COMMENT On May 8, 1995 this sequence version replaced gi:800073.

Contact: Maché R., Clabault G., Harrik H.  
 Université Grenoble Biologie moléculaire et végétale  
 CNRS-CERMO,  
 BP 53 38041 Grenoble, France  
 Email: rmache@grenet.fr  
 Insert Length: 2833 Std Error: 0.00  
 High quality sequence stop: 200.

FEATURES  
 Source location/Qualifiers  
 1..213  
 /organism="Arabidopsis thaliana"  
 /strain="ecotype Columbia C24"  
 /db\_xref="taxon:3702"  
 /clone="GBGF814"  
 /clone\_lib="Grenoble-B"  
 /note="Vector: Lambda ZAP; tissue=flower buds of  
 A.thaliana ecotype Columbia C24; clone\_library=Grenoble-B;  
 Cloning vector: Lambda ZAP."  
 92 a 31 c 21 g 69 t

BASE COUNT

ORIGIN

Query Match 7.5%; Score 17; DB 24; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 AATGATGAATACACAG 104  
 |||||||  
 Db 112 AATGATGAATACACAG 128

RESULT 8  
 W82028 270 bp mRNA EST 12-SEP-1996  
 LOCUS m02e01.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone  
 DEFINITION IMAGE:403896 5', mRNA sequence.  
 ACCESSION W82028  
 NID g1393105  
 VERSION W82028.1 GI:1393105  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 270)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Rhelsting,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 The WashU-HMT Mouse EST Project  
 Unpublished (1996)  
 On May 18, 1995 this sequence version replaced gi:810968.

TITLE Contact: Marra M/Mouse EST Project  
 JOURNAL WashU-HMT Mouse EST Project  
 COMMENT Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:247664  
 Seq primer: -28M13 rev2 from Amersham.

FEATURES  
 Source location/Qualifiers  
 1..270  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:403896"  
 /clone\_lib="Soares mouse p3NMF19.5"

/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTTCACATCTGAAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."

BASE COUNT 61 a 83 c 74 g 52 t

Query Match 7.5%; Score 17; DB 26; Length 270;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 TCCACCAAGACTGAAA 124  
|||||  
Db 144 TCCACCAAGACTGAAA 160

## RESULT 9

AA049828 292 bp mRNA EST 09-SEP-1996  
LOCUS m109a09.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA  
DEFINITION clone IMAGE:475576 5', mRNA sequence.  
ACCESSION AA049828  
NID 91529499  
VERSION AA049828.1 GI:1529499  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 292)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:286320  
Seq primer: -28M13 rev2 from Amersham.

FEATURES  
Source

Location/Qualifiers  
1..292  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_image="IMAGE:475576"  
/clone\_lib="Soares mouse embryo NBME13.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'

TGTTCACATCTGAAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTT  
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
State Univ., from 2 ]; double-stranded cDNA was ligated to  
Eco RI adapters (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT73 vector. Library went through one round of  
normalization, and was constructed by Bento Soares and  
M. Fatima Bonaldo."

BASE COUNT 67 a 88 c 83 g 54 t

Query Match 7.5%; Score 17; DB 27; Length 292;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 TCCACCAAGACTGAAA 124  
|||||  
Db 128 TCCACCAAGACTGAAA 144

## RESULT 10

AA050569 361 bp mRNA EST 09-SEP-1996  
LOCUS m16b12.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA  
DEFINITION clone IMAGE:476255 5', mRNA sequence.  
ACCESSION AA050569  
NID 91530241  
VERSION AA050569.1 GI:1530241  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 361)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On May 8, 1995 this sequence version replaced gi:79561.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:286999  
Seq primer: -28M13 rev2 from Amersham.

FEATURES  
Source

Location/Qualifiers  
1..361  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_image="IMAGE:476255"  
/clone\_lib="Soares mouse embryo NBME13.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCTGAAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTT  
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne



QY	108	TCCACCAAGACCTGAAA	124
Db	409	TCCACCAAGACCTGAAA	425

```

RESULT 15
C63765/c 360 bp mRNA EST 22-SEP-1997
DEFINITION C63765 Yuji Kohara unpublished CDNA Caenorhabditis elegans CDNA
clone YK346es 5', mRNA sequence.
ACCESSION C63765
NID g2422470
VERSION C63765.1 GI:2422470
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota: Metazoa: Nematoda: Secernentea: Rhabditia: Rhabditidae:
AUTHORS Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
Sano, M., Miyata, A. and Nishigaki, A.
Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:693214.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
location/Qualifiers
FEATURES
source
1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/map="5.p13.3-p13.1"
/clone="YK346es"
/clone_id="Yuji Kohara unpublished CDNA"
/note="dev stage-varied, sex-Hermaphrodite male,
tissue.type=whole animal"
BASE COUNT 119 a 59 c 73 g 108 t 1 others
ORIGIN

Query Match 7.5%; Score 17; DB 36; Length 360;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 ATGATGTTCAACACAG 150
|||||
Db 343 ATGATGTTCAACACAG 327

```

Search completed: August 6, 1999, 09:22:56  
 Job time: 6247 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:48 ; Search time 3059.95 Seconds  
(without alignments)  
151.488 Million cell updates/sec

Title: US-09-049-696-12

Perfect score: 235  
Sequence: 1 GACACGACCAATTCCCGAG.....CAGTGTAAAGTGGCGGCTC 235

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	7.7	393	20	T54765	T54765 yb41h12.r1
2	18	7.7	419	26	W35104	W35104 zc70b07.r1
3	18	7.7	215	26	W59838	W59838 zc5700442.M
4	18	7.7	418	38	AA768319	AA768319 oa72e04.s
5	18	7.7	312	40	AA913329	AA913329 o127h10.s
6	18	7.7	334	40	AA936169	AA936169 om06a06.s
7	18	7.7	359	41	A1047469	A1047469 ud66b12.y
8	18	7.7	446	50	A1677052	A1677052 60504807
9	18	7.7	450	50	A1688053	A1688053 wa83b02.x
10	17	7.2	216	22	R53449	R53449 yj79f02.r1
11	17	7.2	147	26	N54133	N54133 yz02h06.r1
12	17	7.2	402	26	W39654	W39654 zc19f05.r1
13	17	7.2	388	26	W51999	W51999 zc37h02.s1
14	17	7.2	326	27	C10507	C10507 c10507 yu1
15	17	7.2	398	28	AA084501	AA084501 zn18e08.s
16	17	7.2	425	29	AA139751	AA139751 mq38h03.x
17	17	7.2	366	29	AA167395	AA167395 zp08a10.s
18	17	7.2	416	29	AA189004	AA189004 zp79g11.s
19	17	7.2	268	30	AA258381	AA258381 zt62b01.s
20	17	7.2	347	31	AA319855	AA319855 EST22135
21	17	7.2	378	32	AA340188	AA340188 EST45359
22	17	7.2	581	33	AA391619	AA391619 LD10668.5
23	17	7.2	340	33	AA406580	AA406580 zv11c05.s
24	17	7.2	456	33	AA434580	AA434580 zw52b10.r
25	17	7.2	443	33	AA436029	AA436029 zu03f12.s
26	17	7.2	431	34	AA456175	AA456175 zxt74c03.r
27	17	7.2	528	34	AA456227	AA456227 zt99a02.r
28	17	7.2	342	34	AA489872	AA489872 ab01c01.r
29	17	7.2	479	35	AA541829	AA541829 vj01f01.r
30	17	7.2	481	35	AA570688	AA570688 nj20c05.s
31	17	7.2	469	36	AA611564	AA611564 v085d05.r
32	17	7.2	629	37	AA667474	AA667474 vv29g08.r
33	17	7.2	308	37	AA667913	AA667913 vv19f08.r
34	17	7.2	711	37	AA699594	AA699594 z142h08.s
35	17	7.2	450	37	AA734088	AA734088 vs19c08.r
36	17	7.2	626	38	AA798468	AA798468 vx75a01.r
37	17	7.2	657	40	AA941117	AA941117 LD25083.5
38	17	7.2	492	40	AA949263	AA949263 LD28088.5
39	17	7.2	652	40	AA950172	AA950172 LD30114.5
40	17	7.2	641	40	AA950203	AA950203 LD30156.5
41	17	7.2	729	40	AA951933	AA951933 LD32845.5
42	17	7.2	416	42	A1079484	A1079484 oz39h01.s
43	17	7.2	460	42	A1082789	A1082789 ox74b11.x
44	17	7.2	508	42	A110896	A110896 MF003D03.
45	17	7.2	207	50	A1702284	A1702284 tz35g04.x

#### ALIGNMENTS

RESULT 1  
T54765/c 393 bp mRNA EST 06-FEB-1995  
LOCUS yb41h12.r1 Stratogene fetal spleen (#937205) Homo sapiens cDNA  
DEFINITION clone IMAGE:73799 5' similar to gb:M91036.rna2  
T54765 HEMOGLOBIN GAMMA-A AND GAMMA-C CHAINS (HUMAN), mRNA sequence.  
ACCESSION T54765  
NID 9656626

VERSION T54765.1 GI:656626  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 393)  
 AUTHORS Hallier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., Dubuque, T., Favellio, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
 TITLE Generation and analysis of 280,000 human expressed sequence tags  
 JOURNAL Genome Res. 6 (9), 807-828 (1996)  
 MEDLINE 97044478  
 COMMENT

TITLE Contact: Wilson RK  
 JOURNAL Washington University School of Medicine  
 MEDLINE 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 COMMENT Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Insert Size: 682  
 High quality sequence stops: 299 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 682 Std Error: 0.00  
 Seq primer: M13rpl  
 High quality sequence stop: 299.

FEATURES  
 source Location/Qualifiers  
 1..393  
 /organism="Homo sapiens"  
 /db\_xref="GDB:495464"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:73799"  
 /clone\_1lb="Stratagene fetal spleen (#937205)"  
 /tissue\_type="fetal"  
 /dev\_stage="fetal"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: spleen; Vector: pBluescript SK-; Site: 1; EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled spleens. Average insert size: 1.0 Kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'";  
 BASE COUNT 93 a 93 c 109 g 97 t 1 others  
 ORIGIN

Query Match 7.7%; Score 18; DB 20; Length 393;  
 Best local Similarity 100.0%; Pred. No. 6.1;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CACGACCAATTCCCGCAG 20  
 Db 384 CACGACCAATTCCCGCAG 367

RESULT 2  
 W55104 419 bp mRNA EST 11-OCT-1996  
 LOCUS zc70007.r1 Soares\_fetal\_heart\_NBHH19W Homo sapiens cDNA clone  
 DEFINITION IMAGE:327637 5', mRNA sequence.  
 W55104  
 ACCESSION W55104  
 NID 91317110  
 VERSION W55104.1 GI:1317110  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 419)

AUTHORS Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
 TITLE The Wash-Merc EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT On Oct 18, 1995 this sequence version replaced g1:1023248.

TITLE Contact: Wilson RK  
 JOURNAL Washington University School of Medicine  
 MEDLINE 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 COMMENT Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 816 Std Error: 0.00  
 Seq primer: MOD.RBG+ET  
 High quality sequence stop: 377.

FEATURES  
 source Location/Qualifiers  
 1..419  
 /organism="Homo sapiens"  
 /db\_xref="GDB:1261839"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:327637"  
 /clone\_1lb="Soares\_fetal\_heart\_NBHH19W"  
 /sex="unknown"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dT) primer (5' TGTACCAATCTGACGTGGAGCGCGCATCTTTTTTTTTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHL19W."  
 BASE COUNT 108 a 88 c 73 g 150 t  
 ORIGIN

Query Match 7.7%; Score 18; DB 26; Length 419;  
 Best local Similarity 100.0%; Pred. No. 6;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCAGCAATTCGCCAGCC 22  
 Db 186 CCAGCAATTCGCCAGCC 203

RESULT 3  
 W59838 215 bp mRNA EST 06-JUN-1996  
 LOCUS zEST00442 Maize Leaf, Stratagene #937005 Zea mays cDNA clone  
 DEFINITION csuh00442 5' end, mRNA sequence.  
 W59838  
 ACCESSION W59838  
 NID 91366576  
 VERSION W59838.1 GI:1366576  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euPhyloPhytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 Poaceae; Zea.  
 REFERENCE 1 (bases 1 to 215)  
 AUTHORS Baysdorfer, C.  
 TITLE The Maize cDNA Program  
 JOURNAL Unpublished (1993)

COMMENT On May 8, 1995 this sequence version replaced gi:799617.

Contact: Baysdorfer C  
California State University  
Dept Biol Sci, California State Univ, Hayward, CA 94542  
Tel: 5108853459  
Fax: 5108854747  
Email: cbaysdor@haywire.csuhayward.edu  
Seq primer: SK.

FEATURES  
source Location/Qualifiers

1..215  
/organism="Homo sapiens"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="csuh00442"  
/clone\_1ib="Maize leaf, Stratagene #937005"  
/note="Vector: Uni-ZAP; Site.1: EcoRI; Site.2: XhoI; mRNA isolated from illuminated leaves and sheaths of 5 week old plant. cDNA directionally cloned into vector."  
BASE COUNT 50 a 40 c 61 g 64 t  
ORIGIN

Query Match 7.7%; Score 18; DB 26; Length 215;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 CAGTGAATGAAAAACAG 115  
|||||  
DB 77 CAGTGAATGAAAAACAG 60

RESULT 4  
AA768319 418 bp mRNA EST 08-FEB-1998  
LOCUS oa72e04.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1317822 3',  
DEFINITION mRNA sequence.  
ACCESSION AA768319  
NID 92819334  
VERSION AA768319.1 GI:2819334  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 418)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
JOURNAL Tumor Gene Index  
COMMENT Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2153497.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html

FEATURES  
source

Insert Length: 891 Std Error: 0.00  
Seq primer: ~40m13 fwd. ET from Amersham  
High quality sequence stop: 411.  
Location/Qualifiers  
1..418  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1317822"

/clone\_1ib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD+),  
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
15'-GTTCACATCTGAGTGGGCGGCCGCTCATTTTCTTTT-  
3'. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 86 a 126 c 92 g 114 t  
ORIGIN

Query Match 7.7%; Score 18; DB 38; Length 418;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 129 ACTGGATATGGAGCAGG 146  
|||||  
DB 106 ACTGGATATGGAGCAGG 123

RESULT 5  
AA913329 312 bp mRNA EST 26-AUG-1998  
LOCUS 0127h10.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
DEFINITION IMAGE:1524739 3', mRNA sequence.  
ACCESSION AA913329  
NID 93052721  
VERSION AA913329.1 GI:3052721  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 312)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
JOURNAL Tumor Gene Index  
COMMENT Unpublished (1997)  
On Nov 13, 1997 this sequence version replaced gi:2559136.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1105 Std Error: 0.00  
Seq primer: ~40m13 fwd. ET from Amersham  
High quality sequence stop: 286.  
Location/Qualifiers

FEATURES  
source

1..312  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1524739"  
/clone\_1ib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker. Site.1: Not I; Site.2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NDHL19W, testis NH1, and B-cell  
NCI\_CGAP\_GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo.  
BASE COUNT 104 a 53 c 53 g 102 t  
ORIGIN

Query Match 7.7%; Score 18; DB 40; Length 312;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 89 TGATTGAATCAGTGAATG 106  
|||||  
Db 52 TGATTGAATCAGTGAATG 35

RESULT 6  
AA936169/c 334 bp mRNA EST 10-JUN-1998  
LOCUS om6a06.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
DEFINITION IMAGE:1540210 3', mRNA sequence.  
ACCESSION AA936169  
NID 93094087  
VERSION AA936169.1 GI:3094087  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 334)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:2044011.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (infoimage.lnl.gov) for further information.  
Insert length: 1094 Std Error: 0.00  
Seq primer: -40m13 fwd: ET from Amersham  
High quality sequence stop: 288.  
Location/Qualifiers

## FEATURES

source

1. 334  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1540210"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung N9H119W, testis N9T, and B-cell  
NCL1CGAP\_GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 111 a 58 c 55 g 110 t  
ORIGIN

Query Match 7.7%; Score 18; DB 40; Length 334;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 89 TGATTGAATCAGTGAATG 106  
|||||

Db 50 TGATTGAATCAGTGAATG 33

RESULT 7  
A1047469/c 359 bp mRNA EST 08-JUL-1998  
LOCUS u66b12.y1 Sugano mouse liver m1a Mus musculus cDNA clone  
DEFINITION IMAGE:1450847 5' similar to gb:M33388.tn1 CYTOCHROME P450  
(HUMAN); gb:M24263 Mouse testosterone 16-alpha-hydroxylase  
(MUSE); mRNA sequence.  
ACCESSION A1047469  
NID 93295756  
VERSION A1047469.1 GI:3295756  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 359)  
AUTHORS Marra,M., Haller,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:639254.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (infoimage.lnl.gov) for further information.  
MGI:924163  
Seq primer: custom primer used  
High quality sequence stop: 281.  
Location/Qualifiers

## FEATURES

source

1. 359  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1450847"  
/clone\_lib="Sugano mouse liver m1a"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: liver; Vector: pME18S-FL3; Site\_1: DraIII  
(CACTGCTG); Site\_2: DraIII (CACCATGCG); 1st strand cDNA  
was primed with an oligo(dt) primer  
[ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was  
ligated to a DraIII adaptor (TGTTCGCTCTGCG), digested  
and cloned into distinct DraIII sites of the pME18S-FL3  
vector (5' site CACTGCTG, 3' site CACCATGCG). XhoI should  
be used to isolate the cDNA insert. Size selection was  
performed to exclude fragments <1.5kb. Library  
constructed by Dr. Sumio Sugano (University of Tokyo  
Institute of Medical Science). Custom primers for  
sequencing: 5' end primer CTTGTGCTTAAAGCTGCG and 3' end  
primer CGACCTGCACTCGACACAC."  
BASE COUNT 79 a 86 c 102 g 92 t  
ORIGIN

Query Match 7.7%; Score 18; DB 41; Length 359;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GACACGACCAATTCGCC 18  
|||||

Db 316 GACACGACCAATTCGCC 239

# RESULT 8

LOCUS A1677052 446 bp mRNA EST 24-MAY-1999  
DEFINITION 605048G07.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays  
cDNA, mRNA sequence.

ACCESSION A1677052

NID 94885932

VERSION A1677052.1 GI:4885932

KEYWORDS EST.

SOURCE Zea mays.

# ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Zea.

1 (bases 1 to 446)

# REFERENCE

## AUTHORS

Maize ESTs from various cDNA libraries sequenced at Stanford University

Unpublished (1999)

JOURNAL On Jun 22, 1998 this sequence version replaced gi:3246799.

# COMMENT

Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605048 row: G column: 07.

# FEATURES

## source

Location/Qualifiers

1..446  
/organism="Zea mays"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/clone\_id="605 - Endosperm cDNA library from Schmidt lab"  
/tissue\_type="pericarp, embryo, and endosperm"  
/dev\_stage="10 days post-pollination"  
/lab\_host="DH5(alpha)"  
/note="Organ: Kernel; Vector: pAD-GA4-2; Site:1: EcoRI;  
Site:2: XhoI; Kernel endosperm cDNA library from Schmidt  
lab"

BASE COUNT 129 a 107 c 108 g 102 t

# ORIGIN

Query Match 7.7%; Score 18; DB 50; Length 446;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 CAATTCAGGGCCAGTG 79

Db 157 CAATTCAGGGCCAGTG 174

# RESULT 9

LOCUS A1688053 450 bp mRNA EST 27-MAY-1999  
DEFINITION w83b02.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2302731 3', mRNA sequence.

ACCESSION A1688053

NID 94899347

VERSION A1688053.1 GI:4899347

# KEYWORDS

## SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 450)

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

# JOURNAL

Tumor Gene Index  
Unpublished (1997)

COMMENT On Mar 10, 1998 this sequence version replaced gi:2948046.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Glbco

High quality sequence stop: 437.

# FEATURES

## source

1..450  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="2302731"  
/clone\_id="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pTRT3D-Pac (Pharmacia) with  
a modified polylinker; Site:1: Not I; Site:2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCI CGAP GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 150 a 84 c 67 g 149 t

# ORIGIN

Query Match 7.7%; Score 18; DB 50; Length 450;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 TGATTGATCAGTGATG 106

Db 52 TGATTGATCAGTGATG 35

# RESULT 10

LOCUS R55449 216 bp mRNA EST 22-MAY-1995  
DEFINITION y179f02.r1 Soares breast 2NBHst Homo sapiens cDNA clone  
IMAGE:154971 5', mRNA sequence.

ACCESSION R55449

NID 9824744

VERSION R55449.1 GI:824744

# KEYWORDS

## SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

# REFERENCE

1 (bases 1 to 216)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Parsons, D., Riklin, L., Rongling, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project

Unpublished (1995)

# TITLE

## JOURNAL

## COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
Insert Size: 766  
High quality sequence stops: 141

Source: IMAGE Consortium, LNL.  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 766 Std Error: 0.00  
Seq primer: M13Rpl  
High quality sequence stop: 141.  
Location/Qualifiers

FEATURES  
source

1. 216  
/organism="Homo sapiens"  
/db\_xref="GDB:557094"  
/db\_xref="taxon:9606"  
/map="12"  
/clone\_1lb="IMAGE:154971"  
/clone\_1lb="Soares breast 2NDBSt"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pRT73D (Pharmacia) with a  
modified polylinker; site:1: Not I; site:2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer (5'  
TGTTACCAATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3'),  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of a modified pRT73 vector (Pharmacia).  
Library went through one round of normalization to a Cot -  
230. Library constructed by Bento Soares and M.Felina  
Bonaldo."

BASE COUNT 83 a 33 c 23 g 74 t 3 others  
ORIGIN

Query Match 7.2%; Score 17; DB 22; Length 216;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 200 ACACGAATGGTAGATAC 216  
|||||  
Db 38 ACACGAATGGTAGATAC 54

RESULT 11  
N54133/c 147 bp mRNA EST 30-JAN-1997  
LOCUS Y202H06.r1 Soares\_multiple\_sclerosis\_2NDBHSP Homo sapiens cDNA  
DEFINITION clone IMAGE:281915 5', mRNA sequence.  
ACCESSION N54133  
NID 91195299  
VERSION N54133.1 GI:1195299  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 147)  
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chisoso,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohlfing,T., Scheinberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,  
Trevisakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.  
and Marra,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
On Nov 21, 1995 this sequence version replaced gi:1069654.  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1176 Std Error: 0.00  
Seq primer: 77  
High quality sequence stop: 103.  
Location/Qualifiers

FEATURES  
source

1. 147  
/organism="Homo sapiens"  
/db\_xref="GDB:3900443"  
/db\_xref="taxon:9606"  
/clone="IMAGE:281915"  
/clone\_1lb="Soares\_multiple\_sclerosis\_2NDBHSP"  
/sex="male"  
/tissue\_type="multiple sclerosis lesions"  
/dev\_stage="Age 46"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pRT73D (Pharmacia) with a modified  
polylinker V-type; phagemid; site:1: Not I; site:2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer (5'  
TGTTCACATCTGAGTGGAGCGCGCCATTTTCTTTTCTTTT 3'),  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pRT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot - 5. Library constructed by Bento  
Soares and M.Felina Bonaldo. RNA from 4 multiple sclerosis  
lesions from one patient was kindly provided by Dr. Kevin  
G. Becker (NINDS/NIH)."

BASE COUNT 52 a 17 c 23 g 53 t 2 others  
ORIGIN

Query Match 7.2%; Score 17; DB 25; Length 147;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 200 ACACGAATGGTAGATAC 216  
|||||  
Db 136 ACACGAATGGTAGATAC 120

RESULT 12  
W39654 402 bp mRNA EST 10-OCT-1996  
LOCUS zc19f05.r1 Soares\_parathyroid\_tumor\_NbHPP Homo sapiens cDNA clone  
DEFINITION IMAGE:322785 5', mRNA sequence.  
ACCESSION W39654  
NID 91321363  
VERSION W39654.1 GI:1321363  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 402)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
The Washu-Merck EST Project  
Unpublished (1995)  
On Nov 29, 1993 this sequence version replaced gi:430054.  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 959 Std Error: 0.00

Seq primer: mob.REGA+ET  
High quality sequence stop: 280.  
FEATURES  
Location/Qualifiers  
source  
1. .402  
/organism="Homo sapiens"  
/db\_xref="GDB:1260443"  
/db\_xref="taxon:9606"  
/clone="IMAGE:322785"  
/issue\_lib="Soares\_parathyroid\_tumor\_NBHPA"  
/tissue\_type="parathyroid tumor"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: parathyroid gland; vector: p773D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer  
15' -  
TGTACCAATCTGAAGTGGAGCGCGCCGACCAATTTTTTTTTTTTTTTTTTTT  
T-3' double-stranded cDNA was size selected, ligated to  
Eco RI adapters (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified p773  
vector (Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
adenomas was kindly provided by Dr. Stephen Marx, National  
Institute of Diabetes and Digestive and Kidney Diseases,  
NIH."

BASE COUNT 133 a 69 c 49 g 145 t 6 others  
ORIGIN

Query Match 7.2%; Score 17; DB 26; Length 402;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 200 ACACGATGTAGATAC 216  
|||||  
DB 48 ACACGATGTAGATAC 64

RESULT 13  
LOCUS W51999 388 bp mRNA EST 11-OCT-1996  
DEFINITION Zc37h02.s1 Soares\_senescent\_fibroblasts\_MNHST Homo sapiens cDNA  
ACCESSION W51999  
NID G1349876  
VERSION W51999.1 GI:1349876  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 388)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucada, T., Le, M., Lennon, G., Matra, M.,  
Parsons, J., Riklin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
On Nov 29, 1993 this sequence version replaced gi:430173.  
TITLE  
JOURNAL  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert length: 2055 Std Error: 0.00  
Seq primer: mob.REGA+ET

High quality sequence stop: 271.  
FEATURES  
Location/Qualifiers  
source  
1. .388  
/organism="Homo sapiens"  
/db\_xref="GDB:1256043"  
/db\_xref="taxon:9606"  
/map="7"  
/clone="IMAGE:324531"  
/issue\_lib="Soares\_senescent\_fibroblasts\_MNHST"  
/tissue\_type="senescent fibroblast"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: p773D (Pharmacia) with a modified  
polylinker V-TYPE; phagemid; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer 15' -  
TGTACCAATCTGAAGTGGAGCGCGCCGACCAATTTTTTTTTTTTTTTTTTTT  
T-3' double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified p773 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 142 a 72 c 53 g 119 t 2 others  
ORIGIN

Query Match 7.2%; Score 17; DB 26; Length 388;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 AGGATTTTCAACCTTA 197  
|||||  
DB 90 AGGATTTTCAACCTTA 106

RESULT 14  
LOCUS C10507 326 bp mRNA EST 28-DEC-1998  
DEFINITION C10507 Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA  
clone yk147d7 3', mRNA sequence.  
ACCESSION C10507  
NID G1535578  
VERSION C10507.1 GI:1535578  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;  
Rhabditina; Rhabditidae; Rhabditidae; Ploederinae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 326)  
Kohara, Y., Motonashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,  
Sano, M., Miyata, A. and Nishigaki, A.  
Expression map of the C. elegans genome  
Unpublished (1996)  
On Jan 24, 1995 this sequence version replaced gi:634356.  
TITLE  
JOURNAL  
COMMENT

Contact: Yuj1 Kohara  
Gene Library Lab  
National Institute of Genetics  
Yata 111, Mishima, Shizuoka 411, Japan  
Tel: 0559-75-0771  
Fax: 0559-75-6240  
Email: yk@nig.ac.jp.  
FEATURES  
Location/Qualifiers  
source  
1. .326  
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(e1489)"  
/db\_xref="taxon:6239"  
/clone="yk147d7"  
/issue\_lib="Yuj1 Kohara unpublished cDNA"  
/note="dev\_stage=varied, sex=Hermaphrodite male,  
tissue\_type=whole animal"

BASE COUNT 74 a 74 c 76 g 95 t 7 others  
ORIGIN

Query Match 7.2%; Score 17; DB 27; Length 326;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 TATGAGCAGGTGCTG 151  
DB 73 TATGAGCAGGTGCTG 89

RESULT 15

AA084501 398 bp mRNA EST 31-JUL-1997  
LOCUS zn18e08.s1 Striatene neuroepithelium NT2RAMI 937234 Homo sapiens  
DEFINITION CDNA clone IMAGE:547814 3', mRNA sequence.  
ACCESSION AA084501  
NID 91626575  
VERSION AA084501.1 GI:1626575  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 398)  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Roehlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Meg, J.,  
Trevaastis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
COMMENT On May 18, 1995 this sequence version replaced gi:811231.

TITLE JOURNAL MEDLINE  
FEATURES SOURCE  
1.398  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:3926330"  
/db\_xref="taxon:9606"  
/map="1:14"  
/clone="IMAGE:547814"  
/clone\_lib="Striatene neuroepithelium NT2RAMI 937234"  
/dev\_stage="Ntera-2/RA+M1 neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2  
(Ntera-2/cl.D1) precursor cells induced with Retinoic  
Acid for 1 week, followed by 3 weeks in mitotic inhibitors  
(Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR  
Vector; -5' adaptor sequence: 5' GAATTCGCGCAGAG 3' -3'  
adaptor sequence: 5' CTCGAGTTCCTTTTTCCTTTT 3'."

BASE COUNT 138 a 65 c 54 g 136 t 5 others  
ORIGIN

Query Match 7.2%; Score 17; DB 28; Length 398;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 AGGATTTTCACACTTA 197  
DB 108 AGGATTTTCACACTTA 124

Search completed: August 6, 1999, 09:22:52  
Job time: 6243 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:13:45 ; Search time 3264.21 Seconds

(without alignments)  
228.960 Million cell updates/sec

Title: US-09-049-696-12

Perfect score: 1 GACACCACCAATTCCCGAG.....CAGGTATAAAGTGGCGCTC 235

Scoring table:

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl:.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pil.\*  
8: gb\_p12.\*  
9: gb\_p1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_sc.\*  
14: gb\_sy.\*  
15: gb\_un.\*  
16: gb\_vl.\*  
17: em\_fun.\*  
18: em\_hum1.\*  
19: em\_hum2.\*  
20: em\_hum3.\*  
21: em\_hum4.\*  
22: em\_in.\*  
23: em\_om.\*  
24: em\_ot.\*  
25: em\_ov.\*  
26: em\_pat.\*  
27: em\_ph.\*  
28: em\_pl.\*  
29: em\_ro.\*  
30: em\_sy.\*  
31: em\_un.\*  
32: em\_vl.\*  
33: gb\_ba1.\*  
34: gb\_ba2.\*  
35: gb\_ba3.\*  
36: gb\_ba4.\*  
37: gb\_ba5.\*  
38: gb\_ba6.\*  
39: gb\_ba7.\*  
40: gb\_ba8.\*  
41: gb\_ba9.\*  
42: gb\_ba10.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	235	100.0	3311	11	AF039400	Homo sapi

2	235	100.0	2826	42	AF127036	AF127036 Homo sapi
3	176.4	75.1	2937	12	AB017156	AB017156 Mus muscu
4	148.2	63.1	35278	11	AF039401	AF039401 Homo sapi
5	116.4	49.5	3022	12	AF047838	AF047838 Mus muscu
6	114.8	48.9	3471	12	AF052746	AF052746 Mus muscu
7	107.4	45.7	3604	9	AB026833	AB026833 Homo sapi
8	105.4	44.9	3415	42	AF043976	AF043976 Homo sapi
9	103.4	44.0	3317	3	AF001261	AF001261 Bos tauru
10	103.4	44.0	3288	3	AF001262	AF001262 Bos tauru
11	103.4	44.0	2820	3	AF001263	AF001263 Bos tauru
12	101.4	43.1	2984	3	BTU36445	BTU36445 Bos tauru
13	65	27.7	878	5	195746	195746 Sequence 8
14	35.6	15.1	3163	2	AF022810	AF022810 Haemophil
15	35.6	15.1	10980	2	U32749	U32749 Haemophilus
16	35.4	15.1	29605	7	AB020747	AB020747 Arabidops
17	34.8	14.8	319000	1	AF000006	AF000006 Pyrococcu
18	34.4	14.6	37384	36	CELE04F6	U28943 Caenorhabdi
19	34.2	14.6	339	1	HPA19415	AF009415 Helicobac
20	33.4	14.2	135793	10	AC001231	AC001231 Genomic s
21	32.6	13.9	339	1	HPA19186	AF009186 Helicobac
22	32.6	13.9	339	1	HPA19194	AF009194 Helicobac
23	32.6	13.9	339	1	HPA19402	AF009402 Helicobac
24	32.6	13.9	339	1	HPA19403	AF009403 Helicobac
25	32.6	13.9	339	1	HPA19409	AF009409 Helicobac
26	32.6	13.9	100000	9	AF000030	AF000030 Homo sapi
27	32.6	13.9	100000	9	AF000134	AF000134 Homo sapi
28	32.6	13.9	100000	9	AF000212	AF000212 Homo sapi
29	32.6	13.9	64990	9	AF000251	AF000251 Homo sapi
30	32.4	13.8	100490	35	AC007776	AC007776 Homo sapi
31	31.6	13.4	235141	11	AC004615	AC004615 Homo sapi
32	31.6	13.4	110384	34	HSU159619	AF078462 Homo sapi
33	31.4	13.4	531	1	TU049440	U49440 Thermococcu
34	31.4	13.4	37968	7	CAC20C1	AL033391 C.albicac
35	31	13.2	339	1	HPA19177	AF009177 Helicobac
36	31	13.2	339	1	HPA19205	AF009205 Helicobac
37	31	13.2	339	1	HPA19388	AF009388 Helicobac
38	31	13.2	339	1	HPA19390	AF009390 Helicobac
39	31	13.2	339	1	HPA19395	AF009395 Helicobac
40	31	13.2	339	1	HPA19397	AF009397 Helicobac
41	31	13.2	339	1	HPA19398	AF009398 Helicobac
42	31	13.2	339	1	HPA19414	AF009414 Helicobac
43	31	13.2	34331	1	MTC165	Z95584 Mycobacteri
44	31	13.2	1579	7	ATP10R2	X95738 A.thaliana
45	31	13.2	762	7	GMENDD552	X69157 G.max ENDD5

#### ALIGNMENTS

RESULT 1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS	AF039400				
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,				
ACCESSION	AF039400				
NID	94009457				
VERSION	AF039400.1	GI:4009457			
KEYWORDS	complete cds.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.				
JOURNAL	Genomic cloning, molecular characterization, and functional				
MEDLINE	analysis of human CLCA1, the first human member of the family of				
REFERENCE	Ca2+-activated Cl- channel proteins				
AUTHORS	Genomics 54 (2), 200-214 (1998)				
TITLE	2 (bases 1 to 3311)				
	Gruber,A.D., Elble,R. and Pauli,B.U.				
	Direct Submission				





```
YKGCPEVLOSOTKASIMFAOHVDSIVECTEONHKNKCAPNKNOKNCNRSTMEVI
RDSDEPKTTPTMTOPNPPTFSLSIQIGRICVLDKSGSMATGRNRLRLNQAQCLPL
LQIVELGSMVMTFDSAAHVQSELIQINSFSDRDLAKRLPAAAGSTISCSGRSA
FTVIRKRYPTDGSSEIVLITDEDMTISCFEVEKQSGAIHTTVALGSPAQAELELSK
MTGILQIVASDOVONNGILIDAFALSSGNGAVSQSALQIESKGLTLONSQMMNGVIY
DSYQKOTLFLITWTTOPOLIMDPSGOKGGEFVDNTPKAYLOIPIGIAKVGTMKY
SLOASORLTLITVSASRNATLPTITVSKTNKSPRPSPPLVYANIROGASPIKRA
SVTALIBSVNKTIVTELLDNGAGADAKDQGVSRFTYTDYDNGRYSVKRABIGVA
ARRRVIPQSGALYIPGWIENDEIOMNPPEPKNDVQHKQVCFSTSSGSEFVAS
DVNAPIDLPFGQITDLNAEIHGSLINTWTAPGDYDGHAKYIIRISTILD
LRDKFNEISLQVNTALLIPKEANSEVEFLKEENITFENGTLDTFLAIQAVDVKSEI
SNIRVSLFIPQTPPTPPSPDETSPACPNIHINSTIGIHLIKIMMWIGELQISIA
"
exon      6667. 6807
           /gene="hCLCA1"
           /number=3
exon      6979. 7126
           /gene="hCLCA1"
           /number=4
exon      9696. 9801
           /gene="hCLCA1"
           /number=5
exon      15454. 15631
           /gene="hCLCA1"
           /number=6
exon      18591. 18809
           /gene="hCLCA1"
           /number=7
exon      19783. 20010
           /gene="hCLCA1"
           /number=8
exon      22253. 22427
           /gene="hCLCA1"
           /number=9
exon      24522. 24628
           /gene="hCLCA1"
           /number=10
exon      26644. 26859
           /gene="hCLCA1"
           /number=11
exon      27447. 27708
           /gene="hCLCA1"
           /number=12
exon      28766. 28936
           /gene="hCLCA1"
           /number=13
exon      31840. 32079
           /gene="hCLCA1"
           /number=14
exon      32919. 33522
           /gene="hCLCA1"
           /number=15
exon      6870 c 7230 g 9933 t
BASE COUNT 11245 a 6870 c 7230 g 9933 t
ORIGIN
Query Match      63.1%; Score 148.2; DB 11; Length 35278;
Best Local Similarity 95.0%; Pred. No. 1.8e-31;
Matches 153; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
RESULT 5
AF047838      3022 bp      mRNA      ROD      26-NOV-1998
LOCUS      AF047838      3022 bp      mRNA      ROD      26-NOV-1998
DEFINITION      Mus musculus calcium-sensitive chloride conductance protein-1
                  (mClCA1) mRNA, complete cds.
ACCESSION      AF047838
VERSION      AF047838.1 GI:3925280
KEYWORDS
SOURCE      house mouse.
ORGANISM      Mus musculus
                Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 3022)
AUTHORS      Gandhi,R., Eblie,R.C., Gruber,A.D., Schreuer,K.D., Ji,H.L.,
                Fuller,C.M. and Pauli,B.U.
TITLE      Molecular and functional characterization of a calcium-sensitive
                chloride channel from mouse lung
JOURNAL      J. Biol. Chem. 273 (48), 32096-32101 (1998)
MEDLINE      99041980
REFERENCE      2 (bases 1 to 3022)
AUTHORS      Gandhi,R., Eblie,R.C., Gruber,A.D. and Pauli,B.U.
TITLE      Direct Submission
JOURNAL      Submitted (12-FEB-1998) Vet. Pathology, Cornell Univ., VMC C4-146,
                Ithaca, NY 14856, USA
FEATURES
    source      Location/Qualifiers
                1..3022
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /tissue_type="whole lung"
                1..3022
                /gene="mClCA1"
                18..2726
                /gene="mClCA1"
                /note="calcium-activated chloride channel; Alternative
                chloride channel to CFTR that may influence severity of
                Cystic Fibrosis in different tissues; mCaCC-1"
                /codon_start=1
                /product="calcium-sensitive chloride conductance
                protein-1"
                /protein_id="AAC79982.1"
                /db_xref="PID:93925281"
                /db_xref="GI:3925281"
                /translation="MWPGLQVLELTFLHLIONTSSVHNSNGEYVIAINPSVE
                DERLIPSEIKENVTQASTYLEFASGGRYFNRIISITLPTNFKSKREYIMPRRESIDKAD
                VIVADPHLQGDPPYTLQYGCGDGRGYIHPTFPLTLDNRLITGPRGRVFWHMAHL
                RMGVFDEYNVDOPFYMSRKNITETRGSTRITGNVNEERNGCVTRACORSTKTL
                YEPKCTFIPDKICTAGASIMEMONLNVVECTEKNVNAEAPNIONMCHNRSTWDVI
                KTSADPQNAPEMREGTEAPPTPEFLSRRVVCLVDKSGMDKEDRLIRMONAEL
                YLQIVKRESMWGLVTFPDSAAHIONLYIKITSSSDYOKITANLPQASGGTSCGHQI
                AGRDATISSDOSTSGSEIVLITDEEDNGISCFEAVRSCHITHTLIGSAAELET
                LDDMTGGLRFYANKDLNSLIDAFERISITSGSVSQALQLESKAFDVRAGAWINGTVP
                LSDVGNDDTFEVIWVWKKEPILIDPKRGKRTTSQDQDLNRSRLQIPGAEFG
                TWYSISGTQSOLITMTVTTRARSPMEPLATAMQSQTAQYPSRAIVYARVSGGL
                PYLGANYTALIEAHGHQVTLLEMDNGADGVYVNDGITYRFDYDNGRYSILKRV
                QAOENKRLSIRQKNKSLYIPGYENKIVLNPRPVOEALATYEDNRRYSGS
                FTVSGAPPDGDHARVFPSPKVTDLAEFFIGDYIHLITATGKVLNDNRARHRIIRMSQ
                HPDLODFNNATLVNASSLIPKEAGKEFEKPEFEKFIKANGIQLYIAIQADNEASL
                TSEVSNIAQAVKLSLSDSDISALGDDISALSMITWGLVIFNSILN"
```

```
BASE COUNT      923 a 710 c 661 g 728 t
ORIGIN
Query Match      49.5%; Score 116.4; DB 12; Length 3022;
Best Local Similarity 69.1%; Pred. No. 1.4e-22;
Matches 159; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
```

Db	1861	CACAGCCACGATACCCTAGCGCGATGATGTGTACGCACGGGTACGCCAAGAATTTTTGGC	1920
OY	63	AATTCTCAGGCCGACGTGTACACGCCCTTGATTTGAATCACTGAATGAAAAACACTTACCTT       	122
Db	1921	TGTTCTGGAGGCCCATGATGCACAGCCCTTCATGAAAGCTGMAATGACATCACTACACCTT       	1980
OY	123	GGAACACTACTGATATATAGGAGAGGTGCTGATGACTACTAGATGACGGATGCTACTACAG       	182
Db	1981	GGAGCTCTGGACAATAGGGGAGGCTGCTGATACGTTTAAATAATGTATGCACTTACACAAG       	2040
OY	183	GTATTTCCAACACTTATGACACAGATGATGATACAGTGTAAAGATGCGCGG       	232
Db	2041	ATACTTACACATTTATCATCTGSAATATAGATACAGCTTAAGAATGCTGT       	2090
RESULT	7		
LOCUS	AB026833		
DEFINITION	Homo sapiens mRNA for chloride channel protein, complete cds.	PRI	26-MAY-1999
ACCESSION	AB026833		
NID	94887600		
VERSION	AB026833.1 GI:4887600		
KEYWORDS	chloride channel protein. Homo sapiens adult corneal epithelium cDNA to mRNA, clone:1Id:lamba ZAP clone:lamba 13802.		
SOURCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (sites) Itoh.R., Kawamoto,S., Kinoshita,S., Kawasaki,S. and Okubo,K. Isolation and characterization of chloride channel in human corneal epithelium Unpublished (1999)		
REFERENCE	2 (bases 1 to 3604) Itoh,R., Kawamoto,S. and Okubo,K. Direct Submission Submitted (28-APR-1999) to the DDBJ/EMBL/Genbank databases. Reiko Itoh, Institute for Molecular and Cellular Biology, Osaka University; Yamada-oka 1-3,, Suita, Osaka 565-0871, Japan (E-mail:reiko@imcb.osaka-u.ac.jp, Tel:81-6-6879-7992, Fax:81-6-6877-1922)		
JOURNAL			
AUTHORS			
TITLE			
FEATURES			
SOURCE	Location/Qualifiers 1..3604 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /clone-"lamba 13802" /clone_1Id-"lamba ZAP" /dev_stage="adult" /map_"1p32" /tissue_type="corneal epithelium" 42..2873 /codon_start=-1 /product="chloride channel protein" /protein_id="BAAF7810.1" /db_xref="PID:d1041578" /db_xref="GI:4887601" /db_xref="PI:4887601" /translation="MTQRSIAGPICNKFTLLVLAASLEPFLGAVQLDNGYNGLIL IATNPENOMLISNTEKMITESFTLENATRKRVEFRNKLITIPATWRANNSKITIK OESEYKANVLTDMYGAGHDPTTLQRRGGCKBEKYIHFPNFLNLNLVANGSGRGK VFPEHANHLRMVGVEDXINDRPYINOQNDIKVTRCSDDITGVPCRCPOPNICIIH SKFEKECPETIVNSTONATASIMFMOSSLSVEFCNASHTHOEAENLONONGCSIRSA DETIDSDFHSFPMNGTELPPPTFSILOAKGVCLVILDSYSKMAEDRLLOOARL AEYTIOWIVEIHFVFGIASDSKGELIAOLHOINSNDRLTVSYLPTVSANKDISI CSULKRGFEVERLUNGKARGSVMLTVTSGGDKLGNCPLVLSGSSTHSHIALSSAP PNTEELSRLTGLGKTFEPVDNISNSMSIDAERSISSGDI.FEQOHIOESTGEVNKP QLNRTYLVNTVGNDETFVLTYWQASGPPEIILDPDGRKYTNNEIFNLFFRTSLMI PGARKGHWTLYTLNTHSLOALKVTYTSRASNSAVPATVEAEVRSDSLFPPVMIM YAWKGGFVTLNATATATATEPETGDPVTRLIDDGAGADVIKNDGISYRFESFAAN GRSLKHYVNHSPSISTPRAISIGSHAMYPGTTRANONIQMARPKSVGRNEERKKNG FSNVSGGSSVLGEPAGPHDVPEPKCIIDLELVKVEEBELTWSVAEGEDFGQAT SYSIRMSKSIQNIOQDFNNAILVNTSRNRNQOAGIRIFLFESSPOISTNGEPHQNGSTG		

[illegible]



```

/db_xref="GI:2623767"
/translation="MWICLVVILFLTHLLPGKSSMVNLINNGYDGVIAINPSVE
DEKLIENIKEMVTEASTYLEHATKRRTYRNVSILLPMTKSKSEYFIFKOSYDAD
VIYANPYLAKYDDPYLQYRGCEKGYHFTFNPLTNFHLTIGSRGFVHEMHL
RMGJFEDYNVDOFYISRNKTIETAROSTHITGINVFKCPGGSCITSLCRDSOTG
LYEAKCFELPKSOTAKESIMFPSLSHTECTEKTHTNENALQNMCKNGSTWD
IMNSVPONTSPMTENMPHPTFESLSKROVCLVLDKSGMSAEDRLFQNMQAE
LVLIOVERKSIVGMWTPDSVAEIONHLRTIDDNVOKITAKIPOVANGTISICGL
KAGFOALIHSDQSTGSEIILLDGEDNETNSCEPDVKKSGATIIHTIALGPSAKELE
TUSMTGGRFENFARKDITGLNMAFSRISSSGSIQOALOLESKALKITGRKVNCTV
PVDSTVGNDFEFVWTWIOKEIYLODPKCKITSDKEDKLTMSARLQIPGIMFT
GTWTSILNNHASQMLVTVTRARSPITPVYATAMSOHTAHPSPVIVAYQSO
GELPLGISVLAIIETEDGHOVLELMDNGAGRDVKNNGIYSRYFDYGNORYSLK
VHAQRNNTARLNRONKMYLVYGVENKILNPPREVDLAKAKIEDPSRLT
SGSFTVSGAPPPGNHPSVPPPSKITTLEAKFKEDYIQLSMTAPGVNLDRKASGSP
MSRSHOVAKLELOLOHOFQ 755 t

```

```

BASE COUNT 940 a 565 c 560 g 755 t
ORIGIN
Query Match 44.0%: Score 103.4: DB 3: Length 2820;
Best Local Similarity 66.2%: Pred. No. 5.6e-19;
Matches 149: Conservative 0: Mismatches 76: Indels 0: Gaps 0:

```

```

OY 11 AATTCGCCAGCCCTCTGTAGTTATGCAAAATATTCGCCAGAGCCTCCCAATTCCTCA 70
DB 2025 ATTATCTAGCCCAATGATTTTATGACAGTCAAGGCTTTTTCCTGACTGCG 2084
OY 71 GGGCCAGTGTACAGCCCTGATTTGATCAGTGAATGGAAGAGTACTTGAAGTAC 130
DB 2085 GATTCAGTGAATGAGCATTAAGAAACGAGATGAGATCAAGTACATTTGAGACTCT 2144
OY 131 TGGATATGAGAGCAGTGTCTACTAAGATGACGCTGTCTACTCAAGTATTTCA 190
DB 2145 GGGACATAGTGTGAGTGTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 2204
OY 191 CAACCTATGACAGCAATGATGATACAGTGTAAAGTGGGGCTC 235
DB 2205 CAGATTCTATGAAATGATGATACAGTTTAAAGATCATGAC 2249

```

```

RESULT 12
LOCUS BTU36445 2984 bp mRNA MAM 09-FEB-1996
DEFINITION Bos taurus calcium-activated chloride channel mRNA, complete cds.
ACCESSION U36445
VERSION U36445.1 GI:1184065
KEYWORDS
SOURCE
ORGANISM
COW.
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.

```

```

REFERENCE
AUTHORS 1 (bases 1 to 2984)
Cunningham, S.A., Awayda, M.S., Buben, J.K., Ismailov, I.I.,
Arrate, M.P., Berdiev, B.K., Benos, D.J., and Fuller, C.M.
Cloning of an epithelial chloride channel from bovine trachea
J. Biol. Chem. 270 (52), 31016-31026 (1995)
MEDLINE 96125078
TITLE 2 (bases 1 to 2984)
AUTHORS Benos, D.J.
REFERENCE Direct Submission
TITLE Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,
JOURNAL University of Alabama at Birmingham, BHSB 706, Birmingham, AL
35294, USA

```

```

FEATURES
Source location/Qualifiers
1..2984
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_type="epithelial cells"
/tissue_type="trachea"
CDS
19..2730

```

```

/note="epithelial chloride channel"
/codon_start=1
/product="calcium-activated chloride channel"
/protein_id="AAC48511.1"
/db_xref="PID:91184066"
/db_xref="GI:1184066"
/translation="MWICLVVILFLTHLLPGKSSMVNLINNGYDGVIAINPSVE
DEKLIENIKEMVTEASTYLEHATKRRTYRNVSILLPMTKSKSEYFIFKOSYDAD
VIYANPYLAKYDDPYLQYRGCEKGYHFTFNPLTNFHLTIGSRGFVHEMHL
RMGJFEDYNVDOFYISRNKTIETAROSTHITGINVFKCPGGSCITSLCRDSOTG
LYEAKCFELPKSOTAKESIMFPSLSHTECTEKTHTNENALQNMCKNGSTWD
IMNSVPONTSPMTENMPHPTFESLSKROVCLVLDKSGMSAEDRLFQNMQAE
LVLIOVERKSIVGMWTPDSVAEIONHLRTIDDNVOKITAKIPOVANGTISICGL
KAGFOALIHSDQSTGSEIILLDGEDNETNSCEPDVKKSGATIIHTIALGPSAKELE
TUSMTGGRFENFARKDITGLNMAFSRISSSGSIQOALOLESKALKITGRKVNCTV
PVDSTVGNDFEFVWTWIOKEIYLODPKCKITSDKEDKLTMSARLQIPGIMFT
GTWTSILNNHASQMLVTVTRARSPITPVYATAMSOHTAHPSPVIVAYQSO
GELPLGISVLAIIETEDGHOVLELMDNGAGRDVKNNGIYSRYFDYGNORYSLK
VHAQRNNTARLNRONKMYLVYGVENKILNPPREVDLAKAKIEDPSRLT
SGSFTVSGAPPPGNHPSVPPPSKITTLEAKFKEDYIQLSMTAPGVNLDRKASGSP
MSRSHOVAKLELOLOHOFQ 755 t

```

```

BASE COUNT 1027 a 583 c 561 g 813 t
ORIGIN
Query Match 43.1%: Score 101.4: DB 3: Length 2984;
Best Local Similarity 65.9%: Pred. No. 2e-18;
Matches 147: Conservative 0: Mismatches 76: Indels 0: Gaps 0:

```

```

OY 11 AATTCGCCAGCCCTCTGTAGTTATGCAAAATATTCGCCAGAGCCTCCCAATTCCTCA 70
DB 1847 ACTACCTAGCCCAATGATTTTATGACAGTCAAGGCTTTTTCCTGACTGCG 1906
OY 71 GGGCCAGTGTACAGCCCTGATTTGATCAGTGAATGGAAGAGTACTTGAAGTAC 130
DB 1907 GATTCAGTGAATGAGCATTAAGAAACGAGATGAGATCAAGTACATTTGAGACTCT 1966
OY 131 TGGATATGAGAGCAGTGTCTACTAAGATGACGCTGTCTACTCAAGTATTTCA 190
DB 1967 GGGACATAGTGTGAGTGTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 2026
OY 191 CAACCTATGACAGCAATGATGATACAGTGTAAAGTGGGGCTC 233
DB 2027 CAGATTCTATGAAATGATGATACAGTTTAAAGATCATGAC 2069

```

```

RESULT 13
LOCUS 195746 878 bp DNA PAT 17-JUL-1998
DEFINITION Sequence 8 from patent US 5733748.
ACCESSION 195746
VERSION 93940216
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 878)
AUTHORS Yu, G. and Rosen, C.
TITLE Colon specific genes and proteins
JOURNAL Patent: US 5733748-A 8 31-MAR-1998;
FEATURES
Source location/Qualifiers
1..878
/organism="unknown"

```

```

BASE COUNT 257 a 179 c 188 g 241 t 13 others
ORIGIN
Query Match 27.7%: Score 65: DB 5: Length 878;
Best Local Similarity 100.0%: Pred. No. 2.4e-06;
Matches 65: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```







Oy	105	TGAAAAACAGTTACTCTGGAACTACGTGGTATGAGCAGGTCGTGCCTACTCAAGA	164
Dd	3528	TGCAGAAATTGAACCTTAATAAGAGTAATATCACAAAAGGGCGATTCTATTAAACA	3469
Oy	155	TGACGCTGTACTACAGATATTCCAACTATGACAGATGGT	210
Dd	3468	TGTGATGTGCTCTTAGTGCTGTCTGTAATTTTAAACCAAAGAT	3423

Search completed: August 5, 1999, 23:13:55  
Job time: 7718 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:13:06 ; Search time 3264.21 Seconds

(without alignments)  
186.091 Million cell updates/sec

Title: US-09-049-696-11

Perfect score: 191  
Sequence: 1 GCGTTGTGTAGTGCACAAAAA.....AACAGAGACACGCAAAATT 191

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_dal:\*  
2: gb\_baz:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr1:\*  
11: gb\_pr2:\*  
12: gb\_pr3:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hvg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hvg1:\*  
35: gb\_hvg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: gb\_in3:\*  
39: em\_baz:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	191	100.0	3311	11	AF039400 Homo sapi

2	191	100.0	2826	42	AF127036	AF127036 Homo sapi
3	135.8	71.1	35278	11	AF039401	AF039401 Homo sapi
4	123.8	64.8	2937	12	AB017156	AB017156 Mus muscu
5	35.8	18.7	3317	3	AF001261	AF001261 Bos tauru
6	35.8	18.7	3288	3	AF001262	AF001262 Bos tauru
7	35.8	18.7	2830	3	AF001263	AF001263 Bos tauru
8	32	16.8	3022	12	AF047838	AF047838 Mus muscu
9	32	16.8	3471	12	AF052746	AF052746 Mus muscu
10	31.4	16.4	131234	9	AF128834	AF128834 Homo DNA s
11	31.2	16.3	196589	34	AF128834	AF128834 Homo sapi
12	31	16.2	2984	3	BTU36445	BTU36445 Bos tauru
13	31	16.2	3604	9	AB026833	AB026833 Homo sapi
14	31	16.2	165729	35	AC007254	AC007254 Homo sapi
15	31	16.2	2048	37	MOSGAAR	MOSGAAR Homo sapi
16	30.8	16.1	36507	10	HS53688	L4406 Aedes aegypt
17	30.8	16.1	129756	10	HS875H10	298882 Human DNA s
18	30.8	16.1	224895	34	HSC1966A12	AL034350 Human DNA
19	30.4	15.9	21764	10	HDM178HYD	AL049542 Homo sapi
20	30.4	15.9	46610	11	HSU34879	M84472 Human 17-be
21	30.4	15.9	10127	11	HSU43572	U34879 Human 17-be
22	30.4	15.9	44943	36	CELC44B7	U43572 Human alpha
23	30.2	15.8	4000	7	NCU51117	U28928 Caenorhabd
24	30.2	15.8	2446	17	EVU77384	U51117 Neurospora
25	29.8	15.6	660	7	RRRNA1965	U77384 Ebola virus
26	29.8	15.6	486	14	G23931	AJ001965 Roemeria
27	29.8	15.6	33803	34	AC006614	G23931 human STS W
28	29.6	15.5	5617	9	AB007976	AC006614 Caenorhab
29	29.6	15.5	131845	34	HS249P5	AB007976 Homo sapi
30	29.6	15.5	115863	34	HS268D13	AL035691 Homo sapi
31	29.6	15.5	69992	37	AC005421	AL023513 Homo sapi
32	29.2	15.3	43015	34	AC000050	AC005421 Drosophil
33	29.2	15.3	200877	35	AC007353	AC000050 Homo sapi
34	29.2	15.3	102308	42	AC005231	AC007353 Homo sapi
35	29	15.2	135636	10	HSJ734C18	AC005231 Homo sapi
36	29	15.2	1621	10	HSSPT1	AL049649 Human DNA
37	29	15.2	263628	34	HSJ784K2	Y08685 H. sapiens m
38	29	15.2	261148	34	HSJ1060K6	AL078588 Homo sapi
39	29	15.2	171336	35	AC007255	AL049649 Homo sapi
40	28.8	15.1	78378	8	AC007260	AC007255 Homo sapi
41	28.8	15.1	135039	11	AC006060	AC007260 Arabidops
42	28.8	15.1	51278	19	AC006301	AC006060 Homo sapi
43	28.6	15.1	94988	34	AC005450	AC006301 Drosophil
44	28.6	15.0	137176	11	AC005017	AC005450 Drosophil
45	28.6	15.0	128143	12	AC003063	AC005017 Homo sapi

## ALIGNMENTS

RESULT 1  
AF039400  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA, complete cds.  
DEFINITION  
ACCESSION AF039400  
NID 94009457  
VERSION AF039400.1 GI:4009457  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS Gruber A.D., Eble, R.C., Ji, H.L., Schreuer, K.D., Fuller, C.M. and Paul, B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca<sup>2+</sup>-activated Cl<sup>-</sup> channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 3311)  
AUTHORS Gruber A.D., Eble, R. and Paul, B. U.  
TITLE Direct Submission

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers  
source

1..3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1P22-p31"  
/tissue\_type="small intestine"  
1..3311  
/gene="hCLCA1"  
352..3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPFKSSVFILILHLEGALSNSLIQINNNGEGIVAIIDPNV  
EDVLIQIKDMVQASLYLEATGRKRYFNVALILPEWTKRADYRPLEYKNA  
DVLVAESTPNDPEYTDQNGNCEKGRHILTPDFLAGKLAEGPGKAFFVEMAH  
LRMGVFEYNDDEKYLNGRIQAVRCAGITGVYKCKOGSGCYTKRGTENKVTGL  
YKGECEVLOSROTEKASIMFAQHVDSIVERCTEONNKNKAPNKNOKCNLSRWETI  
RDSDEFKTTPTQPPNPPTPSLIQIRIVCLYLDKSGSMATGNRLNRNOAGQLEL  
LQYVLSWGVMTFDSAAHVQSELIQINSGSDRTLAKRLPAASGCTSCGLRRA  
FTVIRKRYPTDGEIVLITDGEDNTISCFNEVKOSGAIHTVALGPSAOLELELSK  
MTGLOLYASDOVONNGLIDAFGALSNGAVSORSIQLESGLTLQNSQMMNGTVIY  
DSYVQKDTLFLITWTTPQILMDPSGKOGGVNDRKTMAYLQIPGIAKGVTKY  
SLQASSOTLTLTVSRASNATLPITVTSKTNKDTSPPLVYANRQGAFLVRA  
SVTLAIESVNGKTYTELLDNGAGADTKDGVYSRFTTYDNRGRSVYRAGLVN  
AARRVTPQSGALYIPGWIENDEIOMNPPREIKDDVHKOVCSFSSGGSFVAS  
DVPNAPIDLPFGQITDLKAEIHGSLINLTWAPGDDYDHGTAHKYIRISLID  
LRDFNLSLVNTTALIPKRNSEVFLFKPENITFEENGTLFLAIQAVDVLDKSEI  
SNIRVSLFIPQTPPETPSPDETSAPCPNHINSTIPGIHLIKIMMKWIGELQLSIA

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 191; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 2.5e-51;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTTTGTAGTGACAAAACACCAAAATGCGCTTCCAAATCCAGCAATTGCTAAG 60  
|||||  
DB 1972 GCGTTTGTAGTGACAAAACACCAAAATGCGCTTCCAAATCCAGCAATTGCTAAG 2031  
OY 61 GTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTGCAACCTGACCTGCTACAG 120  
|||||  
DB 2032 GTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTGCAACCTGACCTGCTACAG 2091  
OY 121 TCCCGTGGCTCCCAATGCTACCTGCTCCCAATTACAGTACTTCCAAAAGCAAGAGAC 180  
|||||  
DB 2092 TCCCGTGGCTCCCAATGCTACCTGCTCCCAATTACAGTACTTCCAAAAGCAAGAGAC 2151  
OY 181 ACCAGCAATT 191  
|||||  
DB 2152 ACCAGCAATT 2162

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 2826)  
AUTHORS Agnel,M., Verma,T. and Culouscou,J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel,M. and Culouscou,J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrieres, Rue11-Malmaison 92500, France  
FEATURES location/Qualifiers  
source  
1..2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1..2826  
/gene="CACCL1"  
5..2749  
/gene="CACCL1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"

gene  
CDS

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 191; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 2.5e-51;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTTTGTAGTGACAAAACACCAAAATGCGCTTCCAAATCCAGCAATTGCTAAG 60  
|||||  
DB 1625 GCGTTTGTAGTGACAAAACACCAAAATGCGCTTCCAAATCCAGCAATTGCTAAG 1684  
OY 61 GTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTGCAACCTGACCTGCTACAG 120  
|||||  
DB 1685 GTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTGCAACCTGACCTGCTACAG 1744  
OY 121 TCCCGTGGCTCCCAATGCTACCTGCTCCCAATTACAGTACTTCCAAAAGCAAGAGAC 180  
|||||  
DB 1745 TCCCGTGGCTCCCAATGCTACCTGCTCCCAATTACAGTACTTCCAAAAGCAAGAGAC 1804  
OY 181 ACCAGCAATT 191  
|||||  
DB 1805 ACCAGCAATT 1815

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS AF039401  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.

ACCESSION AF039401  
VERSION 94009459  
KEYWORDS AF039401.1 GI:4009459  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
AUTHORS 1 (bases 1 to 35278)  
Gruber A.D., Elble R.C., Ji H.L., Schreuer K.D., Fuller C.M. and Pauli B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
Gruber A.D., Elble R.C. and Pauli B.U.  
AUTHORS Direct Submission  
TITLE Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
JOURNAL Location/Qualifiers  
FEATURES  
source 1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26659,27447..27708,28766..28936,31840..32079,32919..33310)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,6996..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26659,27447..27708,28766..31840,32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPEKSSVFLIIHLHLEGLANSLSLIOLNNGNEGVIVADPNP  
EDETLIQKDVATASLYLFATGKRFFFKVALLIPEWTKTKADYRPRKTKTKNA  
DVLVAESTPNDDEPTQMGNGCGEKGIEHILTPDIFAGKLAETGPOGKAIVHMAH  
LWMGVDEYNDEKFEYLSNGRIQAVRCISAGITGVVVKCOGSCYTKRCPTNKVTGL  
YKGCDFVLOSROTEKASIMFAQVDSIVFCTEONHNKAEAKNOKCNLSTWELI  
RSEDFKKTTPMTOPTPTFSLQIGORIVGLVDKSGMATGNRLNNOAGOLET  
LOTVELSGVMWTFEDSAHVOSLIIQISGSDRDTAKRLPAASGGCSTISGFSRA  
FTVIRKKYPTDSESVILTDGDNITSGCFNFKKSGALITPAASGGCSTISGFSRA  
MTGLQITASDOVNNGLIDAFALSSGGAVSQSIQESGLILQNSQWNGVTVI  
DSTVGKDTLLFTWTQPPQILLMDPSGKOGFVVDKTKNAVLIQPIGIAVGVWIKY  
SLOASSQTLITLTVTSRASNALPIVTVSKTKNDTSKPSPLVYVANIPOGASPIIRA  
SVTALIESVNGATVLELDNGAGADATDDGVSRFTYTDNGRSVSVKVALGAVN  
AARRVITPOGSAIYIPGWIENDEIOWNPPREINKDDQHOHQVCFPSRSSGGSFVAS  
DVPNAPIDLPFPQGITDNLAEIHGGSILNLTWTAPAGDDYDGTAKKIITRISTILD

1LDKRNESLQVNTALLIPKRNSEEPFLPKPENIFENGTDLFLIAQAVDKVDSKEI  
SNIAVNSLFIPTPTPEPSPDETSAPOPNIHINSTIGIHILKIMMKWIGSLQSLIA  
exon 6667..6807  
/gene="hCLCA1"  
exon 6979..7126  
/gene="hCLCA1"  
exon 9696..9801  
/gene="hCLCA1"  
exon 15454..15631  
/gene="hCLCA1"  
exon 18591..18809  
/gene="hCLCA1"  
exon 19783..20010  
/gene="hCLCA1"  
exon 22253..22427  
/gene="hCLCA1"  
exon 24522..24628  
/gene="hCLCA1"  
exon 26644..26859  
/gene="hCLCA1"  
exon 27447..27708  
/gene="hCLCA1"  
exon 28766..28936  
/gene="hCLCA1"  
exon 31840..32079  
/gene="hCLCA1"  
exon 32919..33522  
/gene="hCLCA1"  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 71.1%; Score 135.8; DB 11; Length 35278;  
Best Local Similarity 98.6%; Pred No.1.5e-33;  
Matches 137; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 53 TTGCTAAGGTGGCACTTGGAAATACAGTCTGCAGCAAGCTCACAAACCTTGACCCCTGA 112  
DB 27439 TTGGTCAGGTGGCACTTGGAAATACAGTCTGCAGCAAGCTCACAAACCTTGACCCCTGA 27498  
QY 113 CTGTACGTCGCCGTGGCTCAATGCTACCCCTGCTTCATTTACAGTGACTTCCAAACCA 172  
DB 27499 CTGTACGTCGCCGTGGCTCAATGCTACCCCTGCTTCATTTACAGTGACTTCCAAACCA 27558  
QY 173 ACAAGGACACGACCAAT 191  
DB 27559 ACAAGGACACGACCAAT 27577  
RESULT 4  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS AB017156  
DEFINITION Mus musculus gob-5 mRNA, complete cds.  
ACCESSION AB017156  
VERSION 93721911  
NID AB017156.1 GI:3721911  
KEYWORDS gob-5.  
SOURCE Mus musculus adult intestine goblet cell cdna to mRNA.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites) Komiyama, Y. and Hirohashi, S. Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice

JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)

MEDLINE 99160866

REFERENCE 2 (bases 1 to 2937) Komiyama, Y. and Hirohashi, S. Direct Submission

JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru Komiyama, EPARO, JST, Hirohashi Cell Configuration Project, 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan (E-mail: tkom@ncsp.jst.go.jp, Tel: 81-298-47-7563, Fax: 81-298-47-5226)

FEATURES

Source Location/Qualifiers  
1..2937  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15..2756  
/gene="gob-5"  
15..2756  
/gene="gob-5"  
/codon\_start=1  
/protein\_id="BA33743.1"  
/db\_xref="PID:d1034712"  
/db\_xref="PID:g3721912"  
/db\_xref="GI:3721912"  
/translation="MESKSPVPLILHLGCVLSSESLIQNNNGEIVAIADHPV  
EDEALQIKDMVTQASPTLEATGKRYFNKVALIPESKAKPEYRPLEFKNA  
DVLVSTSPLENDPEYTHIGACGEGKRIHLTPDLAKRLQYGPDRFVEVMAH  
FRMGVFNENDEKRYLSKGRPOAVRCASATLGRNOVRCOGSCITNGKCVIRVIG  
LYKNCVVPDPHONERKASIMENONINVECTEKNNOAENDONORCLRSMEY  
LOESDEKOTPMAPPAOPAPPSILOIGORTVCYLKSGSMINDRNRNOKSLSEI  
LLQVEGSGWGMTFDSAAYQSLKLNAGADRLIKHLPTVSAGSTISGSLRT  
KFTYIKRRYPTDSEIVLLTGDENITSSCDLVKQSAIHLVAGPAARKEQLS  
AMTGLQYSSDOVQNNGLVDAFALSSGNAIAQHSIQLESRGVNLNNOMNGSVI  
VDSVSGDTLFLITWTHTPTIFIMDPGVEQNGEILDTTKVALIYGPYAKGFEMK  
YSIOASQTLTLVTSRAASATLPIITVTPVNNKTKGFPSPVYVASIRGASPIILR  
ASVATLIESVNGKVTLELLDNGADATKNDGYSRFTAFDANGRSYVIMLGGY  
TSDORAPPRNRRATIDGWTLEDGEVRRNPPRETSYQDKQLCSRTSSGGSVAIV  
VPAAPLPDLPCQITDLKASIQGQNLVNLTPAGDDYDHGRASNTIIMKSTIYD  
LRDHFNTSLQNTTGLIPKEASSEIEFEFEFGMTFNGTIFIAIQVADVSNLSEI  
SNIRASVFIAPQEPPIPEDSTPPCPDISINSTIPGIVHLKIMMKWLGEMQVTLGLH"

BASE COUNT 860 a 718 c 693 g 666 t  
ORIGIN

Query Match 64.8%; Score 123.8; DB 12; Length 2937;  
Best Local Similarity 78.0%; Pred. No. 9.4e-30;  
Matches 149; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1 GCGTTTGTAGTGCACAAAACCCAAATGCGCTTCTCAATCCGACGATTGCTAAG 60  
DB 1638 GGTATTACTAGACACCACTAGGCTGCTCAAGTCCAGGACGCTAAG 1697  
QY 61 GTTGACCTTGAATAGCTGCTGACCAACCTGACCTGCTGACG 120  
DB 1698 GTTGCTTTTGAATATACGATTAAGGACCTACAGCTCACTGCTGACG 1757  
QY 121 TCCCGTGGCTCAATGCTACCTGCTCAATTAAGTACTTCCAAAAGCAAGAC 180  
DB 1758 TCCCGTGGCAAGTGCATACCTGCTTATTAAGTACCCCGGTAGTGAATAAGAC 1817  
QY 181 ACCAGCAATTT 191  
DB 1818 ACAGGGAATTT 1828

RESULT 5

AF001261

LOCUS 3317 bp mRNA MAM 19-NOV-1997  
DEFINITION Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA, complete cds.

ACCESSION AF001261  
NID 92623762

VERSION AF001261.1 GI:2623762

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 3317) Gruber, A.D., Abdel-Ghany, M., Levine, R., Eblle, R.C., Widom, J., Goodwin, A., and Pauli, B.U. Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3317) Eblle, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A., and Pauli, B.U. Direct Submission

JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA

FEATURES

Source Location/Qualifiers  
1..3317  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="1"  
/tissue\_type="lung"  
63..2780  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB61529.1"  
/db\_xref="PID:g2623763"  
/db\_xref="GI:2623763"  
/translation="MVLCLNVILFLTLHLPGMKSSMVLNNYDGIIVAINPSVE  
DEKLEIKEMKLTASTYLLPHATKRYFENNVSLIPMTKSEYIIPROESYDAD  
VIYANPLKKGDDPYLTQVRCGKGYHFTPELLLNPHAYSGRYEVRMHL  
RMGTFEDYNDQPYIRKNTLENTRESTHITGYNVFKCPGSGCTISCRBSQIG  
LYEKCTFLPKSOTAKESIMFSLSVSEFCTEKHNEAPYLQKKCNKNGSTWV  
IMNSVDQNTSPMTENMPHPHPTSLKSQRYVCLVLDGSGMSADRLQMNQAAE  
LYLQVLEKGLSVQWTFDSVAEIQNLIRITDQVYQKLTAKLPYANGSTICRGL  
KAGFOALIHSDOSTGSEIILLDGEDNEINSCPEDEVKRSQAIIHTIAGPSAKETE  
TLNMNGGYRFFANKDITGLTNAFSRISRSRSGSITQAOLESKALKITGRKYNCTV  
PDSYVNDPEFVYWTQKPEYLDQPKKKYKTSFKEDKLNIRARQIQRIAT  
GTWTYSLNNHASSQMLVTVTTRAKSPITPPYATIRHMSQHNHPSPIVIAQVSO  
GLPVLGIVIAIILETDGHQVLELVDNAGADYVNDIYSRFTFYDNGYSLK  
VHQAQRNNTARLNRPONKVLVYVGVLEWNGKILNPREVDDLAKATIEFSRLT  
SGSGFTYSGAPENHPSPSKITDLEAKFEDYIOLSTAGVLDGKANSYII  
RISSEFMDREDFDNATLVNTSNLIPKASKEKPEKPHPRENCTKRYISVQAIN  
EAMNISVSHIYVAIKRIPLPDSYHDLGTRKISEITLAILGLMITSVP"

BASE COUNT 1084 a 676 c 634 g 923 t  
ORIGIN

Query Match 18.7%; Score 35.8; DB 3; Length 3317;  
Best Local Similarity 57.4%; Pred. No. 0.17;  
Matches 89; Conservative 0; Mismatches 57; Indels 9; Gaps 1;

QY 37 CTCGAAATCCAGGACATGCTAAGTTGGACCTGGAATACAGCTCT-----GCAA 87  
DB 1734 CTGCAAAATCTGGATGTGCGAGACAGTACTTGACTTACACCTTCAATAATACAT 1793  
QY 88 GCAAGCTCACAACCTTGACCTGATGCTACGTCCTGCTGCTCAATGCTACCTGCTCT 147  
DB 1794 GCCAGCTCTCAAAATGCTAACAGTACGATGACACCTGAGCAGCAAGTCTACTATACC 1853  
QY 148 CCAATTACAGTACTTCCAAAAGCAGACGACAC 182



DB 1854 CCAGTAATTCGCAACAGCTCAGATGATGACATAC 1888

|||||

RESULT 6

AF001262 3288 bp mRNA MAM 19-NOV-1997

LOCUS Bos taurus clone 2 endothelial adhesion molecule Lu-ECAM-1 mRNA, complete cds.

ACCESSION AF001262

NID 92623764

VERSION AF001262.1 GI:2623764

KEYWORDS

SOURCE Bos taurus.

ORGANISM Bos taurus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 3288)

AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.

TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3288)

AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.

TITLE Direct Submission

JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA

FEATURES

source Location/Qualifiers

1..3288

/organism="Bos taurus"

/db\_xref="taxon:9913"

/cell\_type="aortic endothelial cells"

/clone="2"

/tissue\_type="lung"

59..2443

/note="endothelial adhesion molecule; chloride channel"

/codon\_start=1

/product="Lu-ECAM-1"

/protein\_id="AAB86530.1"

/db\_xref="PID:92623765"

/db\_xref="GI:2623765"

/translation="MVLCNVLTFLHLPLPMKSSMVLINNGYDGIYAIINPSVE DEKLLENKEMVTEASTYLPHATKRKYFRVNSILIPMTWKSSEFIIPKOSYQAD VIVANPYLYGDDPTLYGRGEGEKYIHFTPNFLINNFHISGRGVFHEWAHL RWGIDERNVDDPFYISRNKTEARCTHITGINVVRKCPGSCITSLCRDSOTG LYEACTFLPKRSQAKESIMPELSHSTETECTHTENPARNONKCKNCKSTWDY IINNSVDFOPTSPMTENPPTHTFSLKSKORVCLVDKSGMSMEDLFGMNQAE ILYIOVIEKSGLVGWTFEDVAEIGNHLRITDNDYOKITAKLPQVANGSTICGEL KAGFOAIHSDOSTSGSEIILLTDESDNEINCFEDVKRSGAIHITIALGPAARELE TUSNMTGGRFPAKNDITGLTNAFSRISRSISGSIQOALLESKALKITGRKRVNGT PVDSIVGNDTFEVVWTIOKPEIVLQDPGRKTKTSDPEKDLINRSALQIPGIAET GTWTSILNNAHSSOMLTATYVTRARSPITPIVIAHNSOHTAHPSPMIYAAVSQ GFLPVIGLSVIAIETEDGHVYLETLMONGARDVYKNDGITSRYFTDYGGRSLSK VHAQARNNTARLNROPONKVLVPGVYENGKIILNPPREVKDILAKKIDFSLT SGGSFTVSGAPPGNHPVSFPPSKITDLEAKRKEDYIQLSWTAPGNVLDKRGES"

BASE COUNT 1043 a 707 c 645 g 893 t

ORIGIN

Query Match 18.7%; Score 35.8; DB 3; Length 3288;

Best Local Similarity 57.4%; Pred. No. 0.17;

Matches 89; Conservative 0; Mismatches 57; Indels 9; Gaps 1;

DB 1730 CTCGAATATCCAGGACATGCTAGGCTTGGAATAACGCTCTCTAATAATCAT 1789

|||||

DB 88 GCNAGCTCACAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 147

|||||

DB 1790 GCCAGCTCTCAATATGCTACAGATGACATGACATGACATGACATGACATGACAT 1849

DB 1850 CCAGTAATTCGCAACAGCTCAGATGATGACATAC 1884

|||||

RESULT 7

AF001263 2820 bp mRNA MAM 19-NOV-1997

LOCUS Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA, complete cds.

ACCESSION AF001263

NID 92623766

VERSION AF001263.1 GI:2623766

KEYWORDS

SOURCE Bos taurus.

ORGANISM Bos taurus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 2820)

AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.

TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2820)

AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.

TITLE Direct Submission

JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA

FEATURES

source Location/Qualifiers

1..2820

/organism="Bos taurus"

/db\_xref="taxon:9913"

/cell\_type="aortic endothelial cells"

/clone="3"

/tissue\_type="lung"

194..2656

/note="endothelial adhesion molecule; chloride channel"

/codon\_start=1

/product="Lu-ECAM-1"

/protein\_id="AAB86531.1"

/db\_xref="PID:92623767"

/db\_xref="GI:2623767"

/translation="MVLCNVLTFLHLPLPMKSSMVLINNGYDGIYAIINPSVE DEKLLENKEMVTEASTYLPHATKRKYFRVNSILIPMTWKSSEFIIPKOSYQAD VIVANPYLYGDDPTLYGRGEGEKYIHFTPNFLINNFHISGRGVFHEWAHL RWGIDERNVDDPFYISRNKTEARCTHITGINVVRKCPGSCITSLCRDSOTG LYEACTFLPKRSQAKESIMPELSHSTETECTHTENPARNONKCKNCKSTWDY IINNSVDFOPTSPMTENPPTHTFSLKSKORVCLVDKSGMSMEDLFGMNQAE ILYIOVIEKSGLVGWTFEDVAEIGNHLRITDNDYOKITAKLPQVANGSTICGEL KAGFOAIHSDOSTSGSEIILLTDESDNEINCFEDVKRSGAIHITIALGPAARELE TUSNMTGGRFPAKNDITGLTNAFSRISRSISGSIQOALLESKALKITGRKRVNGT PVDSIVGNDTFEVVWTIOKPEIVLQDPGRKTKTSDPEKDLINRSALQIPGIAET GTWTSILNNAHSSOMLTATYVTRARSPITPIVIAHNSOHTAHPSPMIYAAVSQ GFLPVIGLSVIAIETEDGHVYLETLMONGARDVYKNDGITSRYFTDYGGRSLSK VHAQARNNTARLNROPONKVLVPGVYENGKIILNPPREVKDILAKKIDFSLT SGGSFTVSGAPPGNHPVSFPPSKITDLEAKRKEDYIQLSWTAPGNVLDKRGASFSF MSRESHOVAKVLELOLQHSFO"

BASE COUNT 940 a 565 c 560 g 755 t

ORIGIN

Query Match 18.7%; Score 35.8; DB 3; Length 2820;

Best Local Similarity 57.4%; Pred. No. 0.17;

Matches 89; Conservative 0; Mismatches 57; Indels 9; Gaps 1;

DB 1865 CTCGAATATCCAGGACATGCTAGGCTTGGAATAACGCTCTCTAATAATCAT 1924

[illegible]

RESULT	8
AF047838	
LOCUS	AF047838 3022 bp mRNA
DEFINITION	Mus musculus calcium-sensitive chloride conductance protein-1 (mCLCA1) mRNA, complete cds.
	R0D 26-NOV-1998

NID	93925280
VERSION	AF047838.1
	GI:3925280

SOURCE	ORGANISM
house mouse.	
Mus musculus	

REFERENCE 1 (bases 1 to 3022)  
AUTHORS Gandhi, R., Elble, R. C., Gruber, A. D., Schreuer, K. D., Ji, H. L.

**TITLE** Molecular and functional characterization of a calcium-sensitive chloride channel from mouse lung  
**J. Biol. Chem.** 273 (48), 32096-32101 (1998)

2 (bases 1 to 3022)  
Gandhi, R., Elble, R.C., Gruber, A.D. and Pauli, B.U.

Submitted (12-FEB-1998) Vet. Pathology, Cornell Univ., VMC C4-146  
Ithaca, NY 14856, USA

FEATURES	Location/Qualifiers
source	1..3022
	/organism="Mus musculus"
	/db_xref="taxon:10090"
gene	/tissue_type="whole lung"
	1..3022
	/gene="mcl1a1"
CDS	18..2726

```

/gene="mCAL1"
/note="calcium-activated chloride channel; Alternative
of cystic fibrosis in different tissues; mCAC-1"
/codon_start=1
/product="calcium-sensitive chloride conductance
protein-1"
/protein_id="AAC79982.1"
/db_xref="pid:3925281"
/db_xref="gi:3925281"
/translacion="NWPGLOVLLFLHLHLNTSSSMVHLNSGEGVIVAINSPSE
DELLSIKEMTQASTLYFEASOGYRVERFHPNPLDLNLRISGRVAVHEMAH
VIVADHLOHGDPPITLQIGCCGDRQYHRTISCRITGTVHNCBNGCYTATACRSDKTR
RMGDEYVNDQPFYMSKRNTLEATRCSTRTGTVHNCBNGCYTATACRSDKTR
YKCFKTPIDKIQTAGASIMFMQNLNSVFEETENHNAEALNCRKNSCTWYDYL
KTECFQNPAPMRHGETAPPPPTFSLKSRRRYCVLDKSGMDEKLIMNPADEL
VLTQIVKESMGVLTFEDSAHONLITLITSSDYQKITALPOAAGSISGILG
AQOATISSDOSTSGSEVLTLVDGNGIRGSEFAVRSAGIITHIAGSPAARLELE
LSMTGGTGRFIAKDLNLSLDASFSSSSGSOAOLQESKADVAGINSTVNG
LSDVNGDTEFVITWVAKPELIILODPKCKTTSDPODKRNTISAQLOPGRTEG
TWYSITGKSQILITMTVYTRAPSPMEPLTANHSOSTAOPSRMYIARVSGGF
PVGANTATLIEAHGHOVTLIELMDAGADAVKNDGJITRFTYVHNGRSLKVR
OAKRNTKSLRQKNSLYIPGVENGKIVLNPDPDEALIEATVEDFNVYSGG
FTYSGAPPPDSDHARVPSPKVDGLAEAFIDGDIHLMTWAPGVLDNGRAHRTIIMS
HPLDLOENNAITLVANASSIIRKESAKETPFKPEPEKTIANGIOLYIALQDNAS
TSEVSNIAQAVKLTLSLSDISALGDDISAKISMTIMGLYIINFSILN"

```

BASE COUNT  
ORIGIN

- Query Match 16.8%; Score 32; DB 12; Length 30227

Best Local Similarity 53.68; Pred No. 2.9;  
Matches 90; Conservative 0; Mismatches 75; Indels 3; Gaps 1

18 AAACCAAAATGGCTTACCTCCAAATCCAGCATGTGCTAAGGTGGCACTTGGAATA 77

1667 AAACATCCGGTCTGCTAGACTTCAATATCCGGGCACTGCAGAGACAGGTA

78 CAGTCT---GCAAGCAGAGCTCACAAACCTTGACCCTGACTGTACCGTCCCGGTGGGCCAA 134  
||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
1727 CAGCATCAGGGGTACCAAGCTCAGTTGATTACAATGACAGTAGACCACTGGAGCAAGAAG 1786

135 TGCTACCTGCGCTCCATTTACAGTGACTTCCAAAACGACAAAGACAC 182

Db 1787 TCCACCATGAACCACTCTGCTACTGCTCACATGAGTCAGAGCAC 1834

RESULT 9  
AF052746  
-EFFECTIVE  
DATE 08-APR-1999  
EDM

LOCUS	3471 bp	111876	111876	111876
DEFINITION	Mus musculus	chloride channel	CAC	mRNA, complete cds
ACCESSION	AF052746			
VERSION	AF052746.1			

IND	93580340
VERSION	AF052746.1
KEYWORDS	house mouse.
SOURCE	house mouse.

ORGANISM

REFERENCE  
AUTHORS  
Euthera; Rodentia; Sciurognath; Muridae; Murinae; Mus.  
1 (bases 1 to 3471)  
Romio, L., Musante, L., Cintia, R., Seri, M., Moran, O., Zegarra-Moran, O.

TITLE  
and Galletta, L.J.V.  
Characterization of a murine gene homologous to the bovine Cacc

chloride channel  
Gene 228 (1-2), 181-188 (1999)

MEDLINE	991/3882
REFERENCE	2 (bases 1 to 3471)
AUTHORS	Bomfi, T., Musante, L., Clinti, R., Moran, O., Seri, M. and

AUTHORS	TITLE
KOMILO, E. / MEDANEC, E. / SANCHEZ, J. /	Direct Submission
Galletta, L. J. V.	

Submitted (09-MAR-1998) Lab. di Genetica  
Giannina Gaslini, Genova I-16148, Italy

FEATURES	Location/Qualifiers
source	1..3471

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/db_xref="dbEST:AA245677"
47..2752

```

```

/cocon_start=1
/product="chloride channel Cacc"
/protein_id="AAC35003.1"
/db_xref="PID:g3560547"
/db_xref="GI:3560547"
/translation="MVPGLOVLEFLTHLLQNTSSAVHLNSNGEGVIAINPSEVPE
DERILPIKEMWQASTYLFEASOGQVREYRNTSLVPTMKSREPLYLFEKSEYKAD
VIYADPHLOGDDPYLYLOYGGGDRGOYTHENPNLLDNLRIYGRPRVYEHMALL
RMGVFDYINVDQFPIFSKRKTYLEATRCSTRITGTCVHNHNAEPNLOAKNCRSTWDVYI
YKSCDIPDKITAGASIMEMONLSVSEFTECNHNAEPNLOAKNCRSTWDVYI
TERPDTFONAPPMRGTEAPPEPTFLKSRRVAYCLVDSGMDKEDLLIRNNAALL
ALTOIVKESMWGLVTFDSAAHLONYLLKITSSSYOKITALLPOAAGSITLHGLO
AGFATISDSOSTSGSEIYLLDGEDNGJRSCEFAVSRGAILHTIALGPSAAEELTE
LSDMTGGIRFPAKNDLSILDAFSRISSTSGSVSOALLESKAPDAVAGIMGTVAP
LDSVTGDFEFTVIMWKKREILLQDPKCKYITSPQDDKLNISARGLIIPGTAEFG
TWYTSISGTCSOLITLVTYTRAPSPMEPLLVKNDLAAHNSOIAQYPSRMIVYRVSQGF
PVLCAANTALIALEAHGQVLELMDNGADAVKNDLITRYFTYDHNSRYSLSKVRV
PQOAKNRTRLSLRQAKNSLYIPGVYENGKLYLMDPPNDQVLEIATVADNHFYIIRMS
FTVLGADPDGDHARVPEPSVTLDEAEFGDTIHLTWLIPKVALIDEGAHFYIIRMS
HPLDLOGFNNATLYVNASSLIPKEAGSGTEFFKTRFKRIANGIOLYIALIQADNEAS
TSEVNIAQAVKLTSLSDSISALQADISATICMIMGLTVLENSILN"

```

ORIGIN

Query Match	16.88; Score 32; DB 12; Length 34/1
-------------	-------------------------------------



```

/db.xref="SPTRMBL:P79483"
/translation="MVCCLKPQSSSLAALVTLMVSSRLAFAGDPREFLEKRSRC
HFNGSLATLDRYFHNQDEFLFSDSDVAGTAVTLCGRVAFESRMSQDLLEKRG
RVNVCGRHNVGSEFTVORRHPQVTVKATOPLOHNLVCSGSEFPGSIEVAM
FRNQEERKAGVSTGLIONGDMVTOIYMETEVRSGEVTQOVHEPSVSLTVEMR
ARSESQSKMLSGVGGFVLGLFLFAGLFIYFRNKGSGLOPTEFLS"
27490..27778
/note="Aluub repeat: matches 293..3 of consensus"
repeat_region
28554..28856
/note="Alusg repeat: matches 303..3 of consensus"
repeat_region
28926..29218
/note="Aluy repeat: matches 295..2 of consensus"
repeat_region
30213..30276
/note="4 copies of 16 mer 84 & conserved"
repeat_region
32143..32314
/note="MER20 repeat: matches 36..211 of consensus"
repeat_region
36073..36257
/note="MIR repeat: matches 205..8 of consensus"
repeat_region
37180..37249
/note="MIR repeat: matches 54..123 of consensus"
repeat_region
38158..38331
/note="PIR3 repeat: matches 1352..1523 of consensus"
repeat_region
38218..38629
/note="PIR5 repeat: matches 1813..2212 of consensus"
repeat_region
39339..39467
/note="L1MA7 repeat: matches 909..1036 of consensus"
repeat_region
41372..41664
/note="L1ME1 repeat: matches 269..567 of consensus"
repeat_region
41676..41752
/note="UG1 repeat: matches 3..79 of consensus"
repeat_region
41752..41884
/note="L1P83 repeat: matches 765..897 of consensus"
repeat_region
41890..42220
/note="L1MB8 repeat: matches 559..895 of consensus"
repeat_region
43199..43235
/note="L1MD2 repeat: matches 520..556 of consensus"
repeat_region
44375..44595
/note="MSTA repeat: matches 426..201 of consensus"
repeat_region
45097..45178
/note="2 copies of 41 mer 90 & conserved"
repeat_region
45643..45956
/note="Alusk repeat: matches 302..1 of consensus"
repeat_region
47808..47930
/note="MIR repeat: matches 120..243 of consensus"
repeat_region
49379..49450
/note="match: M28584"
misc_feature
50637..50924
/note="Aluub repeat: matches 296..1 of consensus"
repeat_region
50985..51282
/note="Alusc repeat: matches 1..299 of consensus"
repeat_region
54908..55086
/note="MER20 repeat: matches 30..211 of consensus"
repeat_region
55146..55448
/note="Aluy repeat: matches 301..1 of consensus"
repeat_region
55593..55976
/note="LTR3 repeat: matches 19..412 of consensus"
repeat_region
56141..56212
/note="2 copies of 36 mer 93 & conserved"
prim_transcript
<56386..56682
/note="match: EST T05398 clone HFBW45"
repeat_region
57541..57835
/note="Aluy repeat: matches 295..1 of consensus"
repeat_region
59668..59969
/note="Aluy repeat: matches 1..301 of consensus"
repeat_region
60359..60650
/note="Aluy repeat: matches 1..295 of consensus"
misc_feature
61863..61733
/note="match: M28584"
repeat_region
62917..63209
/note="Aluub repeat: matches 297..3 of consensus"
repeat_region
63439..63580
/note="MIR2 repeat: matches 146..5 of consensus"

```

```

Query Match 16.4%; Score 31.4; DB 9; Length 131234;
Best Local Similarity 48.6%; Pred. No. 5.2;
Matches 86; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1 GCGTTGTGATGACAAAACCAAAAGCCCTACTCCAAATCCAGCATGCTAG 60
DB 35547 GGGATTTTCCTGGACAAACACACACAAAGTATGATTCACACCCCTGCTTAATCCAA 35606
QY 61 GTTGACACTTGGAAATACAGCTTCGACAGCAAGCTCAAAACCTGACCTGACTGACG 120
DB 35607 TGTGATCTCGGAAAGAAATATCATTTTACATATGGCTGATATATTCGCTTGCTG 35666
QY 121 TCCGTCGTCGCAATGCTACCTGCTCCATATACGATGCTTCCAAAGCAAG 177
DB 35667 CACGGGGGATACATGTGACGACTGAAATGAATAAGAACTTCAAACTAAATG 35723

RESULT 11
AF128834 DNA 28-FEB-1999
LOCUS Homo sapiens chromosome 8 clone BAC 57G24 map 8p12, WORKING DRAFT
DEFINITION
SEQUENCE, in unordered pieces.
ACCESSION AF128834
MID 64325380
VERSION AF128834.1 GI:4325380
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 196589)
REFERENCE Schudy,A., Nordstie,G., Schlegelberger,B., Drescher,B., Weber,J.,
AUTHORS Schatevovoy,R., Menzel,U. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Biotechnology, Beutenebergstrasse 11, Jena 07745, Germany
COMMENT 1-25567: contig of 25567 bp 25568-25577: gap of unknown size
2578-103442: contig of 77865 bp 103443-103452: gap of unknown size
103453-126189: contig of 22727 bp 126190-126199: gap of unknown
size 126200-135292: contig of 9093 bp 135293-135302: gap of unknown
size 135303-138810: contig of 3508 bp 138811-138820: gap of unknown
size 138821-163732: contig of 24912 bp 163733-163742: gap of
unknown size 163743-196589: contig of 32847 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
location/Qualifiers
source 1..196589
/organism="Homo sapiens"
/db.xref="taxon:9606"
/chromosome="8"
/map="8p12"
/clone="BAC 57G24"
/complement(2340..2634)
/rpt_family="Alusx"
/evidence=not_experimental
2363..2393
/note="Xpound exon prediction, score = 73% (3%)"
exon
/note="Xpound exon prediction, score = 73% (3%)"
repeat_region
complement(3493..3763)
/rpt_family="Aluuo"
/evidence=not_experimental
complement(4152..4449)
/rpt_family="Alusg"
/evidence=not_experimental
6495..6600
/note="GRAL1, score = 60.000%, comment = good shadow"
/evidence=not_experimental
complement(6532..6596)
/note="GRAL1, score = 79.000%, comment = excellent"
/evidence=not_experimental
complement(7430..7637)
repeat_region

```

repeat\_region /rpt\_family="L1p"  
/evidence-not\_experimental  
complement(7822..7927)  
/note="GRAIL, score = 76.000%, comment = excellent"  
exon /evidence-not\_experimental  
complement(7822..7927)  
/note="MZF, score = 72.8%"  
/evidence-not\_experimental  
complement(7825..8411)  
repeat\_region /rpt\_family="L1MC5"  
/evidence-not\_experimental  
complement(8561..8610)  
/note="MZF, score = 62.5%"  
exon /evidence-not\_experimental  
complement(9188..9251)  
/note="GRAIL, score = 90.000%, comment = excellent"  
/evidence-not\_experimental  
complement(10510..10549)  
/note="GRAIL, score = 46.000%, comment = marginal shadow"  
exon /evidence-not\_experimental  
complement(10510..10549)  
/note="GenScan, score = 1.60%, comment = Initial\_exon 40  
bp frame: 0 phase: 1"  
repeat\_region /evidence-not\_experimental  
complement(11908..12224)  
/rpt\_family="AluSc"  
/evidence-not\_experimental  
complement(12533..12575)  
repeat\_region /rpt\_family="MADE1"  
/evidence-not\_experimental  
complement(12797..12837)  
exon /note="Xpound exon prediction, score = 87% (0%)"  
/evidence-not\_experimental  
complement(13111..13210)  
/note="Xpound exon prediction, score = 96% (0%)"  
exon /evidence-not\_experimental  
complement(13971..14277)  
repeat\_region /rpt\_family="AluSg"  
/evidence-not\_experimental  
complement(15035..15091)  
/note="MZF, score = 54.3%"  
exon /evidence-not\_experimental  
complement(16115..16176)  
/note="MZF, score = 70%"  
exon /evidence-not\_experimental  
complement(16429..16543)  
/note="GRAIL, score = 75.000%, comment = excellent shadow"  
exon /evidence-not\_experimental  
complement(16582..30122)  
/note="GRAIL, score = 75.000%, comment = excellent shadow"  
mRNA join(<16453..16582,30122..30189,42512..42614,44566..44672,  
47446..47576,48340..48546,51365..51506,51895..52141)  
/gene="ndf43"  
exon /note="acc.no. U02326"  
/product="ndf 43"  
/evidence-experimental  
complement(16453..16582)  
/gene="ndf43"  
/number=1  
gene /evidence-experimental  
complement(16453..141967)  
/partial  
/note="neu differentiation factor, mRNA acc.no. U02326"  
repeat\_region /evidence-experimental  
complement(17290..17599)  
/rpt\_family="AluY"  
exon /evidence-not\_experimental  
complement(18678..18980)  
repeat\_region /rpt\_family="AluJb"  
/evidence-not\_experimental  
complement(21299..21609)  
repeat\_region /rpt\_family="AluSp"  
/evidence-not\_experimental  
complement(22000..22123)

repeat\_region /rpt\_family="MER85"  
/evidence-not\_experimental  
complement(22493..22600)  
/rpt\_family="L2"  
repeat\_region /evidence-not\_experimental  
complement(22924..22968)  
/note="homology = 86.70%, score = 24, counts = 5"  
repeat\_region /rpt\_type=tandem  
/rpt\_unit=ttttagta  
/evidence-not\_experimental  
complement(23363..23568)  
/rpt\_family="L2"  
repeat\_region /evidence-not\_experimental  
complement(23559..23753)  
/rpt\_family="L1M3"  
exon /evidence-not\_experimental  
complement(23758..23817)  
/note="homology = 96.70%, score = 26, counts = 2"  
repeat\_region /rpt\_type=tandem  
/rpt\_unit=gatttataactgctatttgtaataatg  
/evidence-not\_experimental  
complement(23818..23920)  
/rpt\_family="L1MA7"  
repeat\_region /evidence-not\_experimental  
complement(25786..25937)  
/rpt\_family="AluSg"  
exon /evidence-not\_experimental  
complement(25937..26085)  
/rpt\_family="FLAM\_C"  
repeat\_region /evidence-not\_experimental  
complement(26145..26224)  
/rpt\_family="L2"  
exon /evidence-not\_experimental  
complement(27210..27295)  
repeat\_region /rpt\_family="MIR"  
/evidence-not\_experimental  
complement(27338..27466)  
repeat\_region /rpt\_family="MER5A"  
/evidence-not\_experimental  
complement(28305..28610)  
exon /rpt\_family="AluSg"  
/evidence-not\_experimental  
complement(30122..30189)  
/gene="ndf43"  
/number=2  
repeat\_region /evidence-experimental  
complement(31956..32142)  
/rpt\_family="MER58A"  
repeat\_region /evidence-not\_experimental  
complement(37223..37266)  
/rpt\_family="MIR"  
exon /evidence-not\_experimental  
complement(37296..37354)  
/rpt\_family="MIR"  
repeat\_region /evidence-not\_experimental  
complement(42512..42614)  
/gene="ndf43"  
/number=3  
repeat\_region /evidence-experimental  
complement(43893..44184)  
/rpt\_family="AluJb"  
exon /evidence-not\_experimental  
complement(44546..44672)  
/gene="ndf43"  
/number=4  
repeat\_region /evidence-experimental  
complement(45829..46028)  
/rpt\_family="MIR"  
repeat\_region /evidence-not\_experimental  
complement(46443..46742)  
/rpt\_family="AluSx"  
exon /evidence-not\_experimental  
complement(47446..47576)

Query Match 16.3%; Score 31.2; DB 34; Length 196589;  
Best Local Similarity 48.8%; Pred. No. 6.1;  
Matches 84; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

OY 6 TGTACTGGCAAAACACCAAAATGGCCCTACCTCCAAATCCAGCATGTGTAAGTTGG 65  
DB 142875 TATATGGCAAAAGAAAAAAGAACTCTAGAGATTCAAATATTAGCAGTCACTATTGA 142934  
OY 66 CACTTGAATATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGCTCAGCTCCG 125  
DB 142935 AGTTAGAAAAGAAATAGAGCTGACAGTGAATCAACAAAAAGACAGAAAGANA 142994  
OY 126 TCGCTCAATGCTACCTGCTCCATTCAGTGAAGTCTCCAAAGACAGACAG 177  
DB 142995 TGGAAATTAAGATTCCCTGATCCATTAGGGGTGATACATTAAAGAAAGAG 143046

RESULT 12  
BTU36445 2984 bp mRNA MAM 09-FEB-1996  
LOCUS Bos taurus calcium-activated chloride channel mRNA, complete cds.  
DEFINITION U36445  
ACCESSION 91184065  
NID U36445.1 GI:1184065  
KEYWORDS  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos  
1 (bases 1 to 2984)  
Cunningham, S.A., Awayda, M.S., Bubien, J.K., Ismailov, I.I.,  
Arrate, M.P., Beldiev, B.K., Benos, D.J., and Fuller, C.M.  
Cloning of an epithelial chloride channel from bovine trachea  
J. Biol. Chem. 270 (52), 31016-31026 (1995)  
96125078  
2 (bases 1 to 2984)  
Benos, D.J.  
Direct Submission  
Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,  
University of Alabama at Birmingham, BHSB 706, Birmingham, AL  
35294, USA  
Location/Qualifiers  
1. 2984  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19..2730  
/note="epithelial chloride channel"  
/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="PID:91184066"  
/db\_xref="GI:1184066"  
/translation="MPERLVILFETLLPDKMSSWNLINNGIDGIVAINSPVE  
DEKLQNIKEWTEASTYLEFPAIKRRIKRVFNVSLIPMTWSKSYLIPKESDQAE  
VIVANPYLKHGDDPYTLOYGRGEGKQYIHTPNFLINLNPITGSRARVHMAHL  
RMGJFEDVNDQPEYISRRNTIEARCTSHITGVIVYKCGGSGITPRCRDSQGL  
YEAKCTFPEKSTARESIMOSLSHVECTEKTENHVEANLONKMKNSKSTWDV  
MNSDFQNTSPMTENPOTFSLISKORVCLYLKSGMSSEDLRFMMNADEL  
FLQIIEKSGVAGWTFPSVAIRNNLTKITDDVYENTITANLQPEANGSTICGGLK  
AGFAIIOSQOSTSGSEIILLDGEDNHSCLIEVKOSVYIHTALGPEAAELET  
LSDMTGGRFANDINGITNAFSRISRSOSITQOITLSKALAIETKRWAGTVP  
VDSITGNTFFVVTWTKPEILLQDPGKRYKTSDEKEDLNHSALRPGJLAETG  
TWTSLNNHASPOLLITVTRAPSPPTPVTAFAHKSOMTAHPSVIVYAYVOSG  
FLPVGINVTAIIETEDGHOYTLDMNGADIIINPPREYKDLAKREIDSESLIS  
HAERANNTRLSLRPOPONKALITYETIENGAKTIIINPPREYKDLAKREIDSESLIS  
GGSTFVSGAPGNHPSVLPNNKIIDLEKFEEDHIOUWTAIPANVLDKGNASTYIRI  
KSFLDLOKDFDNATLVNTSSIKPEAGSDENFEKPPRIENGNTFYIAVOAINEA  
NLTSFVSNIAQAIKFIPEPDESVALGTRISAINALIFALMIISIV"

BASE COUNT 1027 a 583 c 561 g 813 t

ORIGIN

Query Match 16.2%; Score 31; DB 3; Length 2984;  
Best Local Similarity 55.5%; Pred. No. 6;  
Matches 86; Conservative 0; Mismatches 60; Indels 9; Gaps 1;

OY 37 CTCGAATCCAGGACGATTCGTAAGTTGGACACTTGGAATACAGTCT-----GCAA 87  
DB 1687 CTCGAAATACCTGGTATGACAGACAGTACTTGACTTCAATTAATCAT 1746  
OY 88 GCAAGCTCAAAACCTTGACCTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 147  
DB 1747 GCAAGCTCAAAATCTACACGACGACGACGACGACGACGACGACGACGACGAC 1806  
OY 148 CCAATTACAGTACTTCCAAACGAAACGAAACGAC 182  
DB 1807 CCAGTACTGCAACAGCTCACATGAGTCAAATATAC 1841

RESULT 13  
AB026833 3604 bp mRNA PRI 26-MAY-1999  
LOCUS Homo sapiens mRNA for chloride channel protein, complete cds.  
DEFINITION AB026833  
ACCESSION 94887600  
NID AB026833.1 GI:4887600  
VERSION  
KEYWORDS chloride channel protein.  
SOURCE Homo sapiens adult corneal epithelium cDNA to mRNA,  
clone, lib:lambda ZAP clone:lambda 13802.  
Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (sites)  
Itoh, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.  
Isolation and characterization of chloride channel in human corneal  
epithelium  
Unpublished (1999)  
2 (bases 1 to 3604)  
Itoh, R., Kawamoto, S. and Okubo, K.  
Direct Submission  
Submitted (28-APR-1999) to the DDBJ/EMBL/Genbank databases. Reiko  
Itoh, Institute for Molecular and Cellular Biology, Osaka  
University, Yamada-oka 1-3, Suita, Osaka, 565-0871, Japan  
(E-mail:reiko@molb.osaka-u.ac.jp, tel:81-6-6879-7992,  
Fax:81-6-6877-1922)  
Location/Qualifiers  
1. 3604  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="lambda 13802"  
/clone\_id="lambda ZAP"  
/dev\_stage="adult"  
/map="lp32"  
/tissue\_type="corneal epithelium"  
42..2873  
/codon\_start=1  
/product="chloride channel protein"  
/protein\_id="BAA77810.1"  
/db\_xref="PID:d1041578"  
/db\_xref="PID:g4887601"  
/db\_xref="GI:4887601"  
/translation="MTORSTAGPICNLKFTVTLVALSSSELPFGAGVQLODNGYGLL  
IATNPQPEQNLIISNTEKITEAFYFNARKRFRFKILIPATWANNNSKIK  
QESYERANLVVDYVGAHGDPTLOYRGCGEGKYIHTPNFLINDNLNLAGVSSRR  
VEVEKHAHLNMGVDEYINDKPTIYNGINGOIKVRCSDITGIFVCEGCGPOBNCII  
SKTEKECTIYNSONTATASIMOSTLSVVEECNASTNOEAPNLOMOCISRSAM  
DVITDSDHFHSPFMKTEILPPPTBSLVAGDKYVCLVDYSKNAEABRLDLOQA  
AEFLYDIAVEIHFTVGASFSDSGEIQAOLHINSNDRLVLVSLFTVSARKDISI  
CSGKKEFEVEENKNGAYSVMLVYSGDRLGNCPLPVLSSGSTHISALGSSAA  
PNLEISLRGLGKLFEPVDISNSMSMDASRLSSGTDGFOOHIOLESIGENYKPHH  
OLKNTVVDNVTGNDMTLVTWQASGPETILLPDPGRKYTYNNFTINLFRASLMT



BASE COUNT 617 a 410 c 545 g 476 t  
ORIGIN

Query Match 16.2%; Score 31; DB 37; Length 2048;  
Best Local Similarity 59.8%; Pred. No. 5.9;  
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
QY 33 CTACTCCAAATCCCGCATTGCTTAAGGTTGGCACTTGGAATACAGTCTGCCAGCAAG 92  
Db 390 CAACACCCCACTACCAAGCCCTTCGAAACTCGAAATCAGAAATCAGAACTAAA 331  
QY 93 CTCACAAACCTTGACCCCTGACTGTCTAC 119  
Db 330 CTCACTCAGCTGTAGCTCATCTCTAC 304

Search completed: August 5, 1999, 23:13:45  
Job time: 7708 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:47 ; Search time 3059.95 Seconds  
(without alignments)  
123.124 Million cell updates/sec

Title: US-09-049-696-11

Perfect score: 191  
Sequence: 1 GGCCTTGTGATGGACAAAAA.....AACAAAGACACGCAAAATT 191

Scoring table: OLIGO\_NWC

Searched: 2546578 segs, 986266752 residues

Database : EST.\*

1: em\_est1:.\*  
2: em\_est2:.\*  
3: em\_est3:.\*  
4: em\_est4:.\*  
5: em\_est5:.\*  
6: em\_est6:.\*  
7: em\_est7:.\*  
8: em\_est8:.\*  
9: em\_est9:.\*  
10: em\_est10:.\*  
11: em\_est11:.\*  
12: em\_est12:.\*  
13: em\_est13:.\*  
14: em\_est14:.\*  
15: em\_est15:.\*  
16: em\_est16:.\*  
17: em\_est17:.\*  
18: em\_est18:.\*  
19: em\_est19:.\*  
20: em\_est20:.\*  
21: em\_est21:.\*  
22: em\_est22:.\*  
23: em\_est23:.\*  
24: em\_est24:.\*  
25: em\_est25:.\*  
26: em\_est26:.\*  
27: em\_est27:.\*  
28: em\_est28:.\*  
29: em\_est29:.\*  
30: em\_est30:.\*  
31: em\_est31:.\*  
32: em\_est32:.\*  
33: em\_est33:.\*  
34: em\_est34:.\*  
35: em\_est35:.\*  
36: em\_est36:.\*  
37: em\_est37:.\*  
38: em\_est38:.\*  
39: em\_est39:.\*  
40: em\_est40:.\*  
41: em\_est41:.\*  
42: em\_est42:.\*  
43: em\_est43:.\*  
44: em\_est44:.\*  
45: em\_est45:.\*  
46: em\_est46:.\*  
47: em\_est47:.\*  
48: em\_est48:.\*  
49: em\_est49:.\*  
50: em\_est50:.\*  
51: em\_est51:.\*  
52: em\_est52:.\*  
53: em\_est53:.\*

54: em\_est22:.\*  
55: em\_est23:.\*  
56: em\_est24:.\*  
57: em\_est25:.\*  
58: em\_est26:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	9.4	520	27	AA051927	AA051927 SWMFCAS54
2	17	8.9	415	20	T05563	T05563 EST03452 Fe
3	17	8.9	417	20	T35174	T35174 EST81287 Hu
4	17	8.9	543	21	T63273	T63273 yC08d10..s1
5	17	8.9	213	22	R79129	R79129 y186c12..s1
6	17	8.9	327	25	N68142	N68142 y255d11..s1
7	17	8.9	438	25	N68218	N68218 y256d11..s1
8	17	8.9	166	26	W18696	W18696 mcl0a04..r1
9	17	8.9	284	26	W70958	W70958 me2a02..r1
10	17	8.9	323	28	AA097614	AA097614 mo08c01..r
11	17	8.9	629	29	AA166751	AA166751 z085c07..s
12	17	8.9	199	30	AA251352	AA251352 z808c02..s
13	17	8.9	428	33	AA402294	AA402294 zu46d06..r
14	17	8.9	385	33	AA429793	AA429793 zw57d06..r
15	17	8.9	317	34	AA464694	AA464694 zx82a08..s
16	17	8.9	303	34	AA522658	AA522658 n17le07..s
17	17	8.9	644	34	AA531016	AA531016 n107c11..s
18	17	8.9	385	35	AA597584	AA597584 29492 Lam
19	17	8.9	376	36	AA602425	AA602425 no96c10..s
20	17	8.9	376	39	AA831491	AA831491 oc61c01..s
21	17	8.9	313	39	AA834992	AA834992 aj50c04..s
22	17	8.9	413	42	A1099549	A1099549 34275 Lam
23	17	8.9	426	44	A1263934	A1263934 q141a08..x
24	17	8.9	342	44	A1303616	A1303616 u181c12..x
25	17	8.9	504	46	A1435120	A1435120 th95e11..x
26	17	8.9	420	47	A1507361	A1507361 v135b12..x
27	17	8.9	442	49	A1636425	A1636425 tz78c08..x
28	17	8.9	534	50	A1694335	A1694335 wd45f02..x
29	16	8.4	365	20	T02837	T02837 FB11EY Feta
30	16	8.4	432	21	R20266	R20266 NIB1960-SR
31	16	8.4	227	21	T70083	T70083 ycl17903..s1
32	16	8.4	240	21	T70153	T70153 ycl17903..r1
33	16	8.4	362	22	R22833	R22833 y947c03..r1
34	16	8.4	501	22	R54364	R54364 y974h12..r1
35	16	8.4	298	22	R59407	R59407 yh17a04..r1
36	16	8.4	511	23	H29665	H29665 ym58b02..r1
37	16	8.4	366	23	H29709	H29709 ym61f05..r1
38	16	8.4	495	23	R87354	R87354 ym86c01..r1
39	16	8.4	360	24	D75010	D75010 CELK091B6F
40	16	8.4	435	24	H91184	H91184 yu98h04..s1
41	16	8.4	162	25	N56313	N56313 j16051F Hum
42	16	8.4	423	25	N60944	N60944 T9ESTzy1490
43	16	8.4	390	53	HSW002181	A1037848 Homo sapi
44	16	8.4				
45	16	8.4				

ALIGNMENTS

RESULT 1  
AA051927 520 bp mRNA EST 12-DEC-1996  
LOCUS SWMFCAS54SK Brugia malayi microfilaria cDNA (SWM94LS-BMf) Brugia  
DEFINITION MALAYI CDNA clone SWMFCAS54 5', mRNA sequence.  
ACCESSION AA051927  
NID 91537094  
VERSION AA051927.1 GI:1537094

KEYWORDS EST.  
SOURCE Brugia malayi.  
ORGANISM Brugia malayi  
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;  
Filarioidea; Onchocercidae; Brugia.

REFERENCE 1 (bases 1 to 520)  
AUTHORS Williams, S.A.  
TITLE Genes expressed in microfilaria of Brugia malayi  
JOURNAL Unpublished (1995)  
COMMENT

CONTACT: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu  
Seq primer: pBluescript SK.  
Location/Qualifiers

FEATURES  
source  
1..520  
/organism="Brugia malayi"  
/strain="TRS Labs"  
/db\_xref="taxon:6279"  
/clone="SMFCA554"  
/clone\_lib="Brugia malayi microfilaria cDNA  
(SAM94LS-BmMf)"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: lambda Unizap XR; Site\_1: EcoR I; Site\_2:  
Xho I; Lymphatic filarial nematode parasite of humans.  
mRNA was prepared from microfilariae of Brugia malayi  
isolated from jirds and converted to double stranded cDNA  
using reverse transcriptase and oligo(dT) followed by  
RNase H and DNase I. The library had 3.5 x 10E5  
independent recombinants and average insert size was 900  
base pairs. The library was constructed by Lori Saunders.  
The library is available from Dr. S.A. Williams, email  
genome@smith.edu."

BASE COUNT 146 a 59 c 85 g 221 t 9 others

ORIGIN

Query Match 9.4%; Score 18; DB 27; Length 520;  
Best Local Similarity 100.0%; Pred. No. 5.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 TTGGCAGCTGGAATATCA 79  
|||||  
Db 373 TTGGCAGCTGGAATATCA 390

RESULT 2  
LOCUS T05563 415 bp mRNA EST 30-JUN-1993  
DEFINITION EST03452 Fetal brain, Striatogene (cat#936206) Homo sapiens cDNA  
clone HFBC39, mRNA sequence.  
ACCESSION T05563  
NID 9316713  
VERSION T05563.1 GI:316713  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 415)  
AUTHORS Adams, M.D., Kerlavage, A.R., Fields, C., and Venter, J.C.  
TITLE 3,400 expressed sequence tags identify diversity of transcripts  
from human brain  
JOURNAL Nature Genet. 4, 256-267 (1993)  
MEDLINE 93364420  
COMMENT

Contact: Adams, MD  
The Institute for Genomic Research

932 Clopper Road, Gaithersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423  
Email: mdadams@tigr.org  
Seq primer: M13-21.  
Location/Qualifiers

FEATURES  
source  
1..415  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):82263"  
/db\_xref="taxon:9606"  
/clone="HFBC39"  
/clone\_lib="Fetal brain, Striatogene (cat#936206)"  
/note="Vector: LambdaZAP-II; 17-18 wk gestation, female;  
oligo-dT + random primed cDNA synthesis; lambdaZAP-II  
vector, 1.0kb average insert size."

BASE COUNT 139 a 64 c 79 g 128 t 5 others

ORIGIN

Query Match 8.9%; Score 17; DB 20; Length 415;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 ACAACCTTGACCCCTGA 112  
|||||  
Db 292 ACAACCTTGACCCCTGA 308

RESULT 3  
LOCUS T35174 417 bp mRNA EST 06-SEP-1995  
DEFINITION EST8187 Human Prostate gland Homo sapiens cDNA 3' end similar to  
None, mRNA sequence.  
ACCESSION T35174  
NID 9617272  
VERSION T35174.1 GI:617272  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 417)  
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,  
Butt, C.J., Lee, N., Kirness, E.F., Weinstein, K.G., Gocayne, J.D.,  
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.-W.,  
Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,  
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S.M.,  
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkley, P.S.,  
Keller, J.M., Kline, K.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M.,  
Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T.,  
Pellegrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L.,  
Saudek, D.M., Shirley, R., Small, R.V., Spriggs, T.A., Utterback, T.R.,  
Weidman, J.F., Yi, Y., Bednarek, D.P., Cao, L., Cepeda, M.A.,  
Coleman, T.A., Collins, E.-T., Dimke, D., Feng, P., Ferrie, A.,  
Fischer, C., Hastings, G.A., He, W.-W., Hu, J.-S., Greene, J.M.,  
Gruber, J., Hudson, P., Kim, A., Kozak, D.L., Kunsch, C., Ji, H., Li, H.,  
Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.-F., Wing, J., Xu, C.,  
Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A.,  
Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.  
Initial Assessment of Human Gene Diversity and Expression Patterns  
Based Upon 83 Million Basepairs of cDNA Sequence  
Nature 377, 3-174 (1995)  
96026280  
Other ESTs: EST81288 THC10759  
Contact: Venter, JC  
The Institute for Genomic Research  
932 Clopper Rd Gaithersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423  
Email: tdbinfo@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please contact the TIGR Database  
(tdbinfo@tigr.org)

FEATURES Seq primer: M13-21.  
Location/Qualifiers

SOURCE

1. .417  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):106371"  
/db\_xref="taxon:9606"  
/clone\_lib="Human Prostate gland"  
/note="Organ: prostate gland"  
BASE COUNT 119 a 111 c 82 g 105 t  
ORIGIN

Query Match 8.9%; Score 17; DB 20; Length 417;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 CCAATCCAGCATTTG 55  
|||||  
Db 10 CCAATCCAGCATTTG 26

RESULT 4  
T63273 543 bp mRNA EST 17-FEB-1995  
LOCUS T63273  
DEFINITION YC08d10.s1 Stratagene lung (#937210) Homo sapiens cDNA clone  
IMAGE:80083 3' similar to contains Alu repetitive element; mRNA  
sequence.

ACCESSION T63273  
NID 9667138  
VERSION T63273.1 GI:667138  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 543)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Maddis, E., Moore, B., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL MEDLINE  
Genome Res. 6 (9), 807-828 (1996)  
97044478

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 592  
High quality sequence stops: 424 Source: IMAGE Consortium, LNL This  
clone is available royalty-free through LNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.  
Insert Length: 592 Std Error: 0.00  
Seq primer: -21m13  
High quality sequence stop: 424.

FEATURES  
SOURCE

Location/Qualifiers  
1. .543  
/organism="Homo sapiens"  
/db\_xref="GDB:483700"  
/db\_xref="taxon:9606"  
/clone="IMAGE:80083"  
/clone\_lib="Stratagene lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site: 1;  
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. normal lung. Average insert size: 1.0 kb.

Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGACGAC  
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."  
BASE COUNT 134 a 152 c 106 g 137 t 14 others  
ORIGIN

Query Match 8.9%; Score 17; DB 21; Length 543;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 ACCGCCAATCCAGGC 51  
|||||  
Db 67 ACCGCCAATCCAGGC 83

RESULT 5  
R79129 213 bp mRNA EST 09-JUN-1995  
LOCUS R79129  
DEFINITION Y186c12.s1 Soares placenta NB2HP Homo sapiens cDNA clone  
IMAGE:146134 3' mRNA sequence.

ACCESSION R79129  
NID 9855410  
VERSION R79129.1 GI:855410  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 213)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holtman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
JOURNAL  
Unpublished (1995)  
COMMENT On May 8, 1995 this sequence version replaced gi:799900.

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 137  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: Promega -21m13  
High quality sequence stop: 137.  
FEATURES  
SOURCE  
Location/Qualifiers  
1. .213  
/organism="Homo sapiens"  
/db\_xref="GDB:557642"  
/db\_xref="taxon:9606"  
/clone="IMAGE:146134"  
/clone\_lib="Soares placenta NB2HP"  
/sex="female"  
/dev\_stage="placenta obtained at birth (full term)"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: placenta; Vector: pT77TD (pharmacia) with a  
modified polylinker; Site: 1; Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dt) primer [5'  
AACGTGAAGATTCGCGCGCCGCGAGATTTTCTTTTCTTTT 3'],"  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT773 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo.

BASE COUNT 80 a 30 c 29 g 69 t 5 others  
ORIGIN

Query Match 8.9%; Score 17; DB 22; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 GACAAACACCAAAAT 29  
 ||||||||||||||||  
 Db 6 GACAAACACCAAAAT 22

RESULT 6  
 N68142 327 bp mRNA EST 13-MAR-1996  
 LOCUS YZ55611.s1 Morton Fetal Cochlea Homo sapiens cDNA clone  
 DEFINITION IMAGE:286965 3', mRNA sequence.  
 N68142  
 ACCESSION G1224303  
 NID N68142.1 GI:1224303  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Chordata; Craniata; Vertebrata; Mammalia;  
 Eukaryota; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 327)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,  
 Chissee, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
 and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 97044478  
 On Nov 29, 1993 this sequence version replaced gi:502211.

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 On Nov 29, 1993 this sequence version replaced gi:502211.

CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: m13 -40 forward  
 High quality sequence stop: 169.  
 Location/Qualifiers  
 1. 327  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3894082"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:286965"  
 /clone\_lib="Morton Fetal Cochlea"  
 /tissue\_type="cochlea"  
 /dev\_stage="16-22 week fetus"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: ear; Vector: pBluescript SK-; Site: 1: EcoRI;  
 Site: 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned  
 unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.  
 37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP  
 XR Vector. Library constructed by N. Robertson, C. Morton.  
 -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTCTTTTCTTTT 3'

BASE COUNT 115 a 47 c 50 g 111 t 4 others

ORIGIN

Query Match 8.9%; Score 17; DB 25; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 96 ACAACCTTGACCTGA 112  
 ||||||||||||||||  
 Db 160 ACAACCTTGACCTGA 176

RESULT 7  
 N68218 438 bp mRNA EST 13-MAR-1996  
 LOCUS YZ56611.s1 Morton Fetal Cochlea Homo sapiens cDNA clone  
 DEFINITION IMAGE:287061 3', mRNA sequence.  
 N68218  
 ACCESSION G1224379  
 NID N68218.1 GI:1224379  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Chordata; Craniata; Vertebrata; Mammalia;  
 Eukaryota; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 438)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,  
 Chissee, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
 and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 97044478  
 On Apr 14, 1993 this sequence version replaced gi:504450.

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 On Apr 14, 1993 this sequence version replaced gi:504450.

CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: m13 -40 forward  
 High quality sequence stop: 384.  
 Location/Qualifiers  
 1. 438  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3894178"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:287061"  
 /clone\_lib="Morton Fetal Cochlea"  
 /tissue\_type="cochlea"  
 /dev\_stage="16-22 week fetus"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: ear; Vector: pBluescript SK-; Site: 1: EcoRI;  
 Site: 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned  
 unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.  
 37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP  
 XR Vector. Library constructed by N. Robertson, C. Morton.  
 -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTCTTTTCTTTT 3'

BASE COUNT 153 a 67 c 62 g 155 t 1 others

ORIGIN

Query Match 8.9%; Score 17; DB 25; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 96 ACAACCTTGACCTGA 112  
 ||||||||||||||||  
 Db 159 ACAACCTTGACCTGA 175

RESULT 8  
 W18696/c 166 bp mRNA EST 10-SEP-1996  
 LOCUS mcl0a04.r1 Soares mouse p3NMFL1.5 Mus musculus cDNA clone  
 DEFINITION IMAGE:348078 5', mRNA sequence.

ACCESSION W18696  
 NID 91294611  
 VERSION W18696.1 GI:1294611  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 166)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On May 9, 1995 this sequence version replaced gi:802688.

CONTACT: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:219878  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 148.

FEATURES  
 source  
 1..166  
 Location/Qualifiers

/organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:348078"  
 /clone\_1lb="Soares mouse p3NMF19.5"  
 /dev\_stage="19.5 dpc total fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Vector: pT73D (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TCTTACCATCTGAGTGGAGCGCGCGCATTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M.Fatima Bonaldo. RNA was kindly provided by  
 Dr. Minoru Ko (Wayne State University)."

BASE COUNT 54 a 22 c 28 g 62 t  
 ORIGIN

Query Match 8.9%; Score 17; DB 26; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 19;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 GGAATATACAGTCTGCAA 87  
 ||||||||||||||||  
 Db 35 GGAATATACAGTCTGCAA 19

RESULT 9  
 LOCUS W70958 284 bp mRNA EST 17-JUN-1996  
 DEFINITION me2a02.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA  
 clone IMAGE:388394 5', mRNA sequence.  
 ACCESSION W70958  
 NID 91379697  
 VERSION W70958.1 GI:1379697  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 284)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Nov 9, 1995 this sequence version replaced gi:1059668.

CONTACT: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:240226  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 279.

FEATURES  
 source  
 1..284  
 Location/Qualifiers

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:388394"  
 /clone\_1lb="Soares mouse embryo NBME13.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TCTTACCATCTGAGTGGAGCGCGCGCAATTTTCTTTTCTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
 14.5dpc embryos (total RNA provided by Minoru Ko, Wayne  
 State Univ., from 2); double-stranded cDNA was ligated to  
 Eco RI adapters (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 pT73 vector. Library went through one round of  
 normalization, and was constructed by Bento Soares and  
 M.Fatima Bonaldo."

BASE COUNT 78 a 33 c 72 g 101 t  
 ORIGIN

Query Match 8.9%; Score 17; DB 26; Length 284;  
 Best Local Similarity 100.0%; Pred. No. 19;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 GGAATATACAGTCTGCAA 87  
 ||||||||||||||||  
 Db 77 GGAATATACAGTCTGCAA 61

RESULT 10  
 LOCUS AA097614 323 bp mRNA EST 25-OCT-1996  
 DEFINITION mo08c01.r1 Life Tech mouse embryo 10 5dpc 10665016 Mus musculus  
 cDNA clone IMAGE:552960 5' similar to gb:M94623 Mouse (MOUSE);,  
 mRNA sequence.  
 ACCESSION AA097614  
 NID 91643297  
 VERSION AA097614.1 GI:1643297  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 323)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
 TITLE The WashU-HMNI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:13933685.

CONTACT: Marra M/Mouse EST Project  
 WashU-HMNI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:333752  
 Seq primer: -28M13 rev1 from Amersham  
 High quality sequence stop: 206.  
 Location/Qualifiers  
 1..323  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:552960"  
 /clone\_lib="Life Tech mouse embryo 10 5dpc 10665016"  
 /tissue\_type="embryo"  
 /dev\_stage="10.5dpc embryos"  
 /lab\_host="DH10B"  
 /note="Organ: whole embryo; Vector: PCMV-SPORT2; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt. 10.5dpc embryos. PCMV-SPORT2 vector."

BASE COUNT 100 a 100 c 81 g 81 t  
 ORIGIN

Query Match 8.9%; Score 17; DB 28; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 AACAGGACACCGCAA 188  
 Db 85 AACAGGACACCGCAA 101

RESULT 11  
 LOCUS AA166751 629 bp mRNA EST 09-NOV-1997  
 DEFINITION z085c07.s1 Striatogene ovarian cancer (#937219) Homo sapiens cDNA  
 clone IMAGE:593676 3', mRNA sequence.  
 ACCESSION AA166751  
 NID g1745198  
 VERSION AA166751.1 GI:1745198  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 629)  
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,S., Jost,S., Kitzman,D., Kucaba,T., Lacy,M., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 TITLE WashU-NCI human EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1393118.  
 CONTACT: Wilson RK  
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1136 Std Error: 0.00  
 Seq primer: -40M13 fwd. from Amersham  
 High quality sequence stop: 424.  
 Location/Qualifiers  
 1..629  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:593676"  
 /clone\_lib="Striatogene ovarian cancer (#937219)"  
 /sex="female"  
 /dev\_stage="adult, 64 years"  
 /lab\_host="SOUR (kanamycin resistant)"  
 /note="Vector: Bluescript SK-; Site:1: EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Papillary serous carcinoma, isolated from ascites, 64 year old caucasian. Average insert size: 0.8 kb; Uni-Zip XR vector; -5' adaptor sequence: 5' GAATTCGGCAG 3' -3' adaptor sequence: 5' CTCGATTTTCTTTTCTTTT 3'"

BASE COUNT 223 a 115 c 101 g 183 t 7 others  
 ORIGIN

Query Match 8.9%; Score 17; DB 29; Length 629;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 GACAAAAACACCAAAAT 29  
 Db 27 GACAAAAACACCAAAAT 43

RESULT 12  
 LOCUS AA251352 199 bp mRNA EST 13-AUG-1997  
 DEFINITION z508c02.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:684578 3', mRNA sequence.  
 ACCESSION AA251352  
 NID g1886315  
 VERSION AA251352.1 GI:1886315  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 199)  
 NCI\_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On May 8, 1995 this sequence version replaced gi:800272.  
 CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1290 Std Error: 0.00  
 Seq primer: -41M13 fwd. Et from Amersham  
 High quality sequence stop: 160.  
 Location/Qualifiers  
 1..199  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:684578"  
 /clone\_lib="NCI\_CGAP\_GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"

```
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT 81 a 25 c 20 g 73 t

ORIGIN

Query Match 8.9%; Score 17; DB 30; Length 199;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 GACAAAAACACCAAAAT 29  
|||||

Db 27 GACAAAAACACCAAAAT 43

RESULT 13  
AA402294 428 bp mRNA EST 09-NOV-1997  
LOCUS z46d06.r1 Soares ovary tumor NBHOT Homo sapiens cDNA clone  
DEFINITION IMAGE:741035 5', mRNA sequence.  
ACCESSION AA402294  
NID 92056207  
VERSION AA402294.1 GI:2056207  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 428)  
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S.,  
Kritman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
Unpublished (1997)  
JOURNAL  
TITLE WashU-NCI human EST Project  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1406980.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 1379 Std Error: 0.00  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 391.

FEATURES  
source  
1..428  
/organism="Homo sapiens"  
/db\_xref="GDB:5941026"  
/db\_xref="taxon:9606"  
/clone="IMAGE:741035"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B"  
/sex="female"  
/tissue\_type="ovarian tumor"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: ovary; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3'],

```
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."
```

BASE COUNT 95 a 128 c 97 g 108 t

ORIGIN

Query Match 8.9%; Score 17; DB 33; Length 428;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 CCAATCCAGCGCATG 55  
|||||

Db 325 CCAATCCAGCGCATG 341

RESULT 14  
AA429793 385 bp mRNA EST 16-OCT-1997  
LOCUS z457d06.r1 Soares total fetus\_Nb2HF8\_9w Homo sapiens cDNA clone  
DEFINITION IMAGE:774155 5', mRNA sequence.  
ACCESSION AA429793  
NID 92113000  
VERSION AA429793.1 GI:2113000  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 385)  
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S.,  
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,  
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-Merck EST Project 1997  
Unpublished (1997)  
JOURNAL  
TITLE WashU-Merck EST Project 1997  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397639.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 376.

FEATURES  
source  
1..385  
/organism="Homo sapiens"  
/db\_xref="IMAGE:774155"  
/clone="lib=Soares\_total\_fetus\_Nb2HF8\_9w"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from pooled 8-9 week  
(total) fetus material with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 96 a 112 c 85 g 92 t

ORIGIN

Query Match 8.9%; Score 17; DB 33; Length 385;





GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:12:28 ; Search time 3264.21 Seconds

(without alignments)  
223.114 Million cell updates/sec

Title: US-09-049-696-10

Sequence: 229

1 GGCACAGTCGATGCTGACGACG.....AACCTTGACCTGACTGTCA 229

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sc:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_v1:\*  
18: em\_fun:\*  
19: em\_hcg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_v1:\*  
34: gb\_hcg1:\*  
35: gb\_hcg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	229	100.0	3311	11	AF039400	Homo sapi

2	229	100.0	2826	42	AF127036	AF127036 Homo sapi
3	173.6	75.8	35278	11	AF039401	AF039401 Homo sapi
4	161.8	70.7	2937	12	AB017156	AB017156 Homo sapi
5	63.8	27.9	3317	3	AF001261	AF001261 Bos tauru
6	63.8	27.9	3288	3	AF001262	AF001262 Bos tauru
7	63.8	27.9	2820	3	AF001263	AF001263 Bos tauru
8	60.6	26.5	2984	3	BTU36445	BTU36445 Bos tauru
9	53.6	23.4	3022	12	AF047838	AF047838 Mus muscu
10	53.6	23.4	3471	12	AF052746	AF052746 Mus muscu
11	52.6	23.0	3415	42	AF043976	AF043976 Homo sapi
12	35	15.3	22000	35	AC005674	AC005674 Homo sapi
13	34.2	14.9	3604	9	AB026833	AB026833 Homo sapi
14	33.6	14.7	110000	34	CEY38E10.1	Continuation (2 of
15	33.6	14.7	69134	34	CEY57A10.4	Continuation (5 of
16	33.6	14.7	33770	36	CEY15A4	Z81062 Caenorhabdi
17	32.6	14.2	80019	34	ATF25E4	AL050399 Arabidops
18	32.4	14.1	2649	9	HS47669	AJ007669 Homo sapi
19	32.4	14.1	2290	11	HSU70310	U70310 Human DNA r
20	32	14.0	186684	35	AC007040	AC007040 Homo sapi
21	31.6	13.8	4624	7	DBPMA1	X73901 D.Dicouclata
22	31.6	13.8	5641	9	AB014609	AB014609 Homo sapi
23	31.6	13.8	4639	42	AF134838	AF134838 Homo sapi
24	31.2	13.6	90143	34	HSAC001232	AC001232 Homo sapi
25	31	13.5	36583	1	SC5H1	AL049863 Streptomy
26	31	13.5	129914	11	AC003070	AC003070 Homo sapi
27	31	13.5	115863	34	HS268D13	AL023513 Homo sapi
28	30.6	13.4	161652	11	AC004774	AC004774 Homo sapi
29	30.6	13.4	54336	11	HUMXPDI1	L47234 Homo sapien
30	30.2	13.2	107600	8	ATF16J13	AL049638 Arabidops
31	30.2	13.2	146213	11	AC004953	AC004953 Homo sapi
32	30.2	13.2	169765	11	AC006075	AC006075 Homo sapi
33	30.2	13.2	22016	17	PCPFRS40	X99448 Feline cali
34	30.2	13.2	323792	34	AC004965	AC004965 Homo sapi
35	30.2	13.2	163794	35	AC007050	AC007050 Hordeum v
36	30	13.1	162405	17	AC002352	AC002352 Homo sapi
37	30	13.1	10644	17	DENSTRA	M14931 Dengue viru
38	30	13.1	231329	34	AC002353	AC002353 Homo sapi
39	30	13.1	138445	34	AC002979	AC002979 Homo sapi
40	30	13.1	206539	34	HS1099D15	AL035456 Homo sapi
41	30	13.1	24064	36	CEFO9B9	Z49887 Caenorhabdi
42	30	13.1	2048	37	MOSGAAR	L44606 Aedes aegypt
43	29.8	13.0	200952	7	ATFC6	Z97341 Arabidops
44	29.8	13.0	38482	36	CEYF6A10	U28736 Caenorhabdi
45	29.8	13.0	37005	36	DMC33C11	AL035331 Drosophila

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	15-DEC-1998
LOCUS	AF039400	3311 bp	15-DEC-1998	
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hCICL1) mRNA, complete cds.			
ACCESSION	AF039400			
NID	94009457			
VERSION	AF039400.1	GI:4009457		
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
REFERENCE	1 (bases 1 to 3311)			
AUTHORS	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.			
TITLE	Genomic cloning, molecular characterization, and functional analysis of human CICL1, the first human member of the family of Ca2+-activated Cl--channel proteins			
JOURNAL	Genomics 54 (2), 200-214 (1998)			
MEDLINE	99047526			
REFERENCE	2 (bases 1 to 3311)			
AUTHORS	Gruber,A.D., Elble,R. and Pauli,B.U.			
TITLE	Direct Submission			

JOURNML Submitted (19-Dec-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPFKSSVFILILHLEGALSNSLIQLNNNGEIVVAIDPNV  
EDELIIQIKDMVQASLYLEATGKRFYFNVALILPEWTKRADYRPLETYKNA  
DVLVAESTPNDPEYTEMONGCEKGERIHLPEDFAGKLAEGYRGAFLVEMAH  
LRMGVFEYNDDEKFIYLSNGRIQAVRSAGITGVVYKCOGSGCYTRCFNFTGL  
YKGCDFELOSROTEKASIMFAOHVDSIVERCTEONHNRKAPNKNOKNRLRSWEVI  
RDSDFKTTPTQPPNPTFSLQIGRIYCLVLDSSGMATGRNLRNLQAQLEL  
LQYVLEGSWMVMTFDSAHVQSELIOINSGBDYLAKRPPAASGTSICSLRSA  
FTVIRKRPIDGSEIVLITDGEDNTISCFENYKQSGAIHTVALGSPAQLELBSK  
MTGILQIYASDOVONNGLIDAFGALSNGNAVSOISLEKGLTLNOSOMNGTVY  
DSTYKQDLFLITWTTPQIILMDPSGOKGVVVKNTKMAVLIQIGIAKMTWY  
SLOASQTLTLTVSRASNATLPIRYTSKNTKPSPLVYVYANRGAASPLRA  
SVTLASVNGKTYTLELDNGAGADATKDGYSRYFTYDTNGRSYVRALAGVNA  
AARRVITPOOSGALYIPGWIENDELIONPPEIKNDVQKQVCSFTSSGGSFVAS  
DVPNAPLPDLFPGQITDLKAEIHGSLINLWTAPGDYDGHYAHKTIIRISTILD  
LRDFENSLQVNTTALIPKEANSEVFLFKPENTTFENGTDLFTAIQAVDVKSEI  
SNIRAVSLFIPQTPPETPSPDETSAFCPNHINSTIPGIHLIKMKMIGELDSIA  
"

gene  
CDS

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 229; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 2,3e-68;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCACAGTGCATCTGGACAGCACCGTGGGAAGACACTTTTCTTATCACTGGACA 60  
|||||  
DB 1861 GGCACAGTGCATCTGGACAGCACCGTGGGAAGACACTTTTCTTATCACTGGACA 1920  
OY 61 ACGCAGCTCCCAAAATCTCTGGGATCCCAAGTGGACAGCAAGAGTGGCTTTGTA 120  
|||||  
DB 1921 ACGCAGCTCCCAAAATCTCTGGGATCCCAAGTGGACAGCAAGAGTGGCTTTGTA 1980  
OY 121 GTGACAAAAACACCAAAATGCGCTACTCTCAAAATCCAGCATTTGTAAGTTGGCACT 180  
|||||  
DB 1981 GTGACAAAAACACCAAAATGCGCTACTCTCAAAATCCAGCATTTGTAAGTTGGCACT 2040  
OY 181 TGAATAATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGCA 229  
|||||  
DB 2041 TGAATAATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGCA 2089

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
AUTHORS Agnel,M., Verma,T. and Culouscou,J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel

JOURNML Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel,M. and Culouscou,J.-M.  
TITLE Direct Submission  
JOURNML Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrieres, Reuil-Malmaison 92500, France  
FEATURES  
source  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="CACCL1"  
5. .2749  
/note="CACCL1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PID:9458469"  
/db\_xref="GI:4585469"  
/translation="MGPFKSSVFILILHLEGALSNSLIQLNNNGEIVVAIDPNV  
EDELIIQIKDMVQASLYLEATGKRFYFNVALILPEWTKRADYRPLETYKNA  
DVLVAESTPNDPEYTEMONGCEKGERIHLPEDFAGKLAEGYRGAFLVEMAH  
LRMGVFEYNDDEKFIYLSNGRIQAVRSAGITGVVYKCOGSGCYTRCFNFTGL  
YKGCDFELOSROTEKASIMFAOHVDSIVERCTEONHNRKAPNKNOKNRLRSWEVI  
RDSDFKTTPTQPPNPTFSLQIGRIYCLVLDSSGMATGRNLRNLQAQLEL  
LQYVLEGSWMVMTFDSAHVQSELIOINSGBDYLAKRPPAASGTSICSLRSA  
FTVIRKRPIDGSEIVLITDGEDNTISCFENYKQSGAIHTVALGSPAQLELBSK  
MTGILQIYASDOVONNGLIDAFGALSNGNAVSOISLEKGLTLNOSOMNGTVY  
DSTYKQDLFLITWTTPQIILMDPSGOKGVVVKNTKMAVLIQIGIAKMTWY  
SLOASQTLTLTVSRASNATLPIRYTSKNTKPSPLVYVYANRGAASPLRA  
SVTLASVNGKTYTLELDNGAGADATKDGYSRYFTYDTNGRSYVRALAGVNA  
AARRVITPOOSGALYIPGWIENDELIONPPEIKNDVQKQVCSFTSSGGSFVAS  
DVPNAPLPDLFPGQITDLKAEIHGSLINLWTAPGDYDGHYAHKTIIRISTILD  
LRDFENSLQVNTTALIPKEANSEVFLFKPENTTFENGTDLFTAIQAVDVKSEI  
SNIRAVSLFIPQTPPETPSPDETSAFCPNHINSTIPGIHLIKMKMIGELDSIA  
"

gene  
CDS

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 229; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 2,2e-68;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCACAGTGCATCTGGACAGCACCGTGGGAAGACACTTTTCTTATCACTGGACA 60  
|||||  
DB 1514 GGCACAGTGCATCTGGACAGCACCGTGGGAAGACACTTTTCTTATCACTGGACA 1573  
OY 61 ACGCAGCTCCCAAAATCTCTGGGATCCCAAGTGGACAGCAAGAGTGGCTTTGTA 120  
|||||  
DB 1574 ACGCAGCTCCCAAAATCTCTGGGATCCCAAGTGGACAGCAAGAGTGGCTTTGTA 1633  
OY 121 GTGACAAAAACACCAAAATGCGCTACTCTCAAAATCCAGCATTTGTAAGTTGGCACT 180  
|||||  
DB 1634 GTGACAAAAACACCAAAATGCGCTACTCTCAAAATCCAGCATTTGTAAGTTGGCACT 1693  
OY 181 TGAATAATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGCA 229  
|||||  
DB 1694 TGAATAATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGCA 1742

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS AF039401  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.

ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota: Metazoa: Chordata: Cranialata: Vertebrata: Mammalia;  
1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CIC1, the first human member of the family of Ca<sup>2+</sup>-activated Cl<sup>-</sup> channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source 1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCIC1"  
1..33522  
/gene="hCIC1"  
767..1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCIC1"  
/note="putative"  
1618..1708  
/gene="hCIC1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCIC1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCIC1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,6996..9801,15454..15631,18591..18809,19783..20010,22253..24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)  
/gene="hCIC1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="GI:4009460"  
/translation="MGPFKSVFLLHLLEGALSNSLIQLNNNGYEGIVADPNP  
EETFLIOIKMVTQASLYLFEATGRFAEFKVAIILPEWTAKDYPKAEYKNA  
DYLVAESTPPGNDPEPTTEOMGNCGEKEIHLTPPIAKKLAIEIGPOKATVHEKAH  
LWGVFDEYNDEKFEYLSNGRIQAVKCSAGITGVNVRKCGGSCYTKRCTFNKVTGL  
YKSGCFYVLOSROTETKASIMFAQHVDSIVEFCTEONHNEAPNKQKONCLSTWEVI  
RSEDEKFTTPMTOTPPNPTFSLIOGRIVLVDKSGSMATGNRLNNOAGOLF  
LOTVELGKVVWVTFDSAAHVOSELIQINSQSDROTALRLPAAGSGTSGISGLSA  
FTVIRKKYPTGSEIILLDGEDNTISGCFNFKVSGALIIHYVAIGSPASAOLEELISK  
MTGGLIYASDOYONNGLIDAFGALSNGGANVSORSIQESGILTIQONQWNGYIV  
DSYVGDIPLITWTQPPQIILLMDPSGKQGGFVVDKNTKNAVLIPGIAVGVTKY  
SLQASSQTLITLVTSASNAATLPITVITVTSKTKNDKSPSPVAVYANIQGASPIIRA  
SVTALIESVNGKTVLELDNGAGADATDDCVSYFETTYDNGSYVAVAGLVGN  
AARRVYPOGSGALYPGWIENDEIQMNPREINDDYOHQOVCSPTSGSGSFVAS  
DVPNAPIPLPFPQGITDLNAILHGSLINLTWTAAGDYDGHGTAHKYIIRISTILD

LDKFNESLOVNTALIPKEANSEVFLFKPENITFENGTLDFIAIQAVDKLSEI  
SNIAVSLFIPQTPPETPSPBETSAPCPNHININSTIGIHILTKIMMKIGSLQSLIA  
".  
6667..6807  
/gene="hCIC1"  
/number=3  
6979..7126  
/gene="hCIC1"  
/number=4  
9696..9801  
/gene="hCIC1"  
/number=5  
15454..15631  
/gene="hCIC1"  
/number=6  
18591..18809  
/gene="hCIC1"  
/number=7  
19783..20010  
/gene="hCIC1"  
/number=8  
22253..22427  
/gene="hCIC1"  
/number=9  
24522..24628  
/gene="hCIC1"  
/number=10  
26644..26859  
/gene="hCIC1"  
/number=11  
27447..27708  
/gene="hCIC1"  
/number=12  
28766..28936  
/gene="hCIC1"  
/number=13  
31840..32079  
/gene="hCIC1"  
/number=14  
32919..33522  
/gene="hCIC1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 75.8%; Score 173.6; DB 11; Length 35278;  
Best local similarity 97.8%; Pred. No. 3.8e-49;  
Matches 176; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GGCACAGTATCGTGGACAGACCGTGGGAAAGACACTTTGTTTATCAGCTGACA 60  
|||||  
DB 26689 GGCACAGTATCGTGGACAGACCGTGGGAAAGACACTTTGTTTATCAGCTGACA 26748  
QY 61 AGCAGAGCTCCCAATTCCTCTCTGGAATCCAGTGGACAGAAAGAGTGGCTTGT 120  
|||||  
DB 26749 AGCAGAGCTCCCAATTCCTCTCTGGAATCCAGTGGACAGAAAGAGTGGCTTGT 26808  
QY 121 GTGACAAAAACACAAATGGCTTACTCTCAATCCAGGCAATTGCTAAGTTGGCACT 180  
|||||  
DB 26809 GTGACAAAAACACAAATGGCTTACTCTCAATCCAGGCAATTGCTAAGTATGAGAT 26868  
RESULT 4  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS AB017156  
DEFINITION AB017156  
ACCESSION AB017156  
VERSION 93721911  
KEYWORDS GOB-5.  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (sites)  
Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice  
JOURNAL  
Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
MEDLINE  
99160866  
REFERENCE  
2 (bases 1 to 2937)  
Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
Submitted (26-AUG-1998) to the DBJ/EMBL/GenBank databases. Tohru Komiyama, ERATO, JST, Hirohashi Cell Configuration Project, 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan  
(E-mail: tkomihcp.jst.go.jp, Tel: 81-298-47-7563, Fax: 81-298-47-5226)  
FEATURES  
source  
1. 2937  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15..2756  
/gene="gob-5"  
15..2756  
/gene="gob-5"  
/codon\_start=1  
/protein\_id="BAA33743.1"  
/db\_xref="PID:01034712"  
/db\_xref="PID:83721912"  
/db\_xref="GI:3721912"  
/translation="MESLSPVFLILHLLEGVLSSELIQNNNGYEGIVAIIDHDVPEDEALIQIKDMVTOASPYLFEATGKRYFNVALIIPESWAKAREYRPLETEKMA DVLVSTSPLENDPEYLGAGCGEGKRIHLTPFLAGKLTQPGDPTFEHMAH FRMGVEYNNDEKPYLSKGPQAVRCSAATGKNQVRGCGSCITGKCVIDVATG LYKNCVFPVPHQNEKASIMENQINSVCEKFNQNOAPNDQNRCLRSWEV IOESDEKOTTPMRAQPPAPTFSLQIQIRYCVLVDLSGSKLNDRLNRNQAQSLR LQIVEDGSGWMTFDSAAVYVSELKQNSGADDLIKHLPTVASAGTISCSGLRT AFTVIAKKRYPDGESEIVLTDGEDNTISSCFDLVQSGAIIHTVALGPAARKEEOL KMTGLOIYSSDOVQNGVLVDFAALSSGNAIAQHSIQLESRGVNLQNNMNGSVI VDSYGRDITFLITWTTPRTFIDPDSVEONGFIIDTTKVALIYVGTAKYGFEMK XSIQASSOTLITVTSRAASATIPITPTVNNKMGKPPSPVAYASIRGASPIILR ASVALIESVNGKITYTTELIDNGAGADTKNDGYSRFTAPDANGRTSYKIMALGV TSDRRAAPKPNRMYIDGWIEDGEVRNPPPEPISYQDKOLCFSSSGSEVAIN VPAAPIDPLEPPCOITDLKASIOGQNLVNLTPAGDQYDHGRASNTIIRMSISIVD LRDHNTSLQVNTTGLIPKEASSEIEFEELGNTGFIQVADVDSNLKSEI SNIAVSVEFIPRAQEPRIPEPDSPPCPDISINSTIGIHVLKIMKWLGEQVTLGLH"

BASE COUNT 860 a 718 c 693 g 666 t  
ORIGIN  
Query Match 70.7%; Score 161.8; DB 12; Length 2937;  
Best Local Similarity 81.7%; Pred. No. 38-45; Indels 0; Gaps 0;  
Matches 187; Conservative 0; Mismatches 42;  
QY 1 GGCACAGTATCGTGGACGACACCGTGGAGAGACACTTGTGTTATACCTGGACA 60  
DB 1527 GGCTCAGTATCGTGGACGACACTGCTGGAGAGACACTTGTGTTATACCTGGACA 1586  
QY 61 AGCGACGCTCCCAATCTCTCTGCGATCCAGTGGACAGAGAAAGTGGCTGTA 120  
DB 1587 AGCGATCCTCCACATATTTATCTGGGATCCAGGCGAGTGGAAACAAATGGTTTATA 1646  
QY 121 GTGACAAAAACCAAAATGGCTACTCTCAATCCAGAGATTGCTAAGTTGGCACT 180  
DB 1647 GTGACACCAACCACTAAGTGGCTACTCTCAAGTCCAGGACGGCTAAGTTGGCTTT 1706  
QY 181 TGGAAATACAGTCTGACGACGACAAACCTTGAACCTGACTCTCA 229  
DB 1707 TGGAAATACAGTCTCAAGGAGCTCAGAGACTCTCACTTGACTCTCA 1755

RESULT 5  
AF001261  
LOCUS  
DEFINITION  
Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA, complete cds.  
ACCESSION  
AF001261  
NID  
92623762  
VERSION  
AF001261.1 GI:2623762  
KEYWORDS  
SOURCE  
ORGANISM  
Bos taurus.  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovine; Bos.  
REFERENCE  
1 (bases 1 to 3317)  
Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.  
Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel  
JOURNAL  
Unpublished  
2 (bases 1 to 3317)  
Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.  
Direct Submission  
TITLE  
Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA  
FEATURES  
source  
Location/Qualifiers  
1. 3317  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="1"  
/tissue\_type="lung"  
63..2780  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB6529.1"  
/db\_xref="PID:92623763"  
/db\_xref="GI:2623763"  
/translation="MVCLNVIIFLTLHLPLGKSSNVNLINNGYDGVIAINPSVE DEKLEIKMKVTEASTYLHFAKRYEYFNVALIPMTKRSSEYIPROESYDQAD VIVANPELTKMGDDPYTLQYRGCGKGYIHETPNFLTNNEHAYSGVREHMAH RMYKEDVYNDQPYISRKNTIATKTRCTSTIGTINVERKCPGSCITSLCRDSQIG LYKACFTPLRSKQSTAKESIMFMSLSVTEFCEKTHNTEARLQKMGKNGSTMY IIMNSVDQNTSPMTENMPPTHPTFSLKRSQVAVCLDLSGSKSADRLQMQAAE LYLQVIEKSLQVMTFDSAAVYVSELKQNSGADDLIKHLPTVASAGTISCSGLRT AFTVIAKKRYPDGESEIVLTDGEDNTISSCFDLVQSGAIIHTVALGPAARKEEOL KMTGLOIYSSDOVQNGVLVDFAALSSGNAIAQHSIQLESRGVNLQNNMNGSVI VDSYGRDITFLITWTTPRTFIDPDSVEONGFIIDTTKVALIYVGTAKYGFEMK XSIQASSOTLITVTSRAASATIPITPTVNNKMGKPPSPVAYASIRGASPIILR ASVALIESVNGKITYTTELIDNGAGADTKNDGYSRFTAPDANGRTSYKIMALGV TSDRRAAPKPNRMYIDGWIEDGEVRNPPPEPISYQDKOLCFSSSGSEVAIN VPAAPIDPLEPPCOITDLKASIOGQNLVNLTPAGDQYDHGRASNTIIRMSISIVD LRDHNTSLQVNTTGLIPKEASSEIEFEELGNTGFIQVADVDSNLKSEI SNIAVSVEFIPRAQEPRIPEPDSPPCPDISINSTIGIHVLKIMKWLGEQVTLGLH"

BASE COUNT 1084 a 676 c 634 g 923 t  
ORIGIN  
Query Match 27.9%; Score 63.8; DB 3; Length 3317;  
Best Local Similarity 60.9%; Pred. No. 1.6e-11; Indels 9; Gaps 1;  
Matches 126; Conservative 0; Mismatches 72;  
QY 1 GGCACAGTATCGTGGACGACACCGTGGAGAGACACTTGTGTTATACCTGGACA 60  
DB 1578 GGCACAGTATCGTGGACGACAGTATGGAATGACACTTCTTGTGTGTCATGACA 1637  
QY 61 AGCGACGCTCCCAATCTCTCTGCGATCCAGTGGACAGAGAAAGTGGCTGTA 111  
DB 1638 ATACAAAAACCAAGAAATGTTCTCCCAAGTCCAAAGAGAAAGAAATATAAAACCTGGAT 1697  
QY 112 GGCTTTAGTGGACAAAAACCAAAATGGCTACTCTCAATCCAGGACTTGTCTAAG 171





```
RMGFEVNDOPFYMRSKNTIETRCSTRITGTGNVYHNCERNGCYTRARBRDSTRL
YERKCTFIPIBKITOTAGASINIFEMONINSYVEFCETREKHNNAAPILQNMCRBRTWYI
KISADPOMAPMRKTEAPPPPTSLKSRRVCLVDKSGMDKEDRLRMQOAEVL
YLQIVKESMWGLVTFDSNAHIONLITITSSDYOKITANLPQASGGSICHGLQ
AGFOAITSDDOSTSGSEIYLLDGEDNGIRSCFEAVRSQAIHTITAGSARELET
LSMTGGLRFEYANKDLNSLIDAFSRISSTGSYSOQALQESKAFVBARAWINGTVP
LSDTVGNDFFEVITMVKRPEIILDPKCKKXTTSPDOKLNRASARLOIPGTAEFG
TWYITSTGTSQILMTVTTTRASPMEPLATAHMSQSTAOQPSRMIYARYSQOGL
PVLGAVNTALIEAHEGHQVTLMDNGAGADTVKNDGITYRTFDTHGNGRYSLSKRV
OQAKNTRLSLRQKNKSLYIPGVENGKIVLNPDPVOEALAEVDEDRYTSKGS
FTVSGAPPDGDHARVPPSKVLDLEAFEGIDYIHLTWAPKYLMDNGRAHRYIIRMSQ
HPDLQEDENNATLVNASSLIPREAGSKETFEKPEFETRIANGIOIYIAIQADNEASTL
TSEVSNIAQVKTLSLEDSISALGDISAISMTIMGLTIFNSILN"

BASE COUNT      923 a      710 c      661 g      728 t
ORIGIN

Query Match      23.4%; Score 53.6; DB 12; Length 3022;
Best Local Similarity 58.5%; Pred. No. 5.1e-08;
Matches 117; Conservative 0; Mismatches 74; Indels 9; Gaps 1;

QY 1 GGCACAGTGAATCGTGACAGCACCCTGGGAAAGACACTTGTTCCTATCACCCTGGACA 60
    || ||||| ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 1530 GGTACAGTACTCTGACAGACGCTGGCAAGACAGCTTCTTGTATCACCCTGGATG 1589

QY 61 ACCGAGCCTCCCAATCTCTCTGGGATCCAGTGACAGAAAGCA-----AGGT 111
    || ||||| ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 1590 GTAAAAAAGCCAGAAATCATCTTCAAGATCCAAAGAAAATAATATACACCTCAGAT 1649

QY 112 GGGTTTGTAGTGACAAAACACCAAAATGGCCTACCTCCAAATCCAGGATGCTAAG 171
    || ||||| ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 1650 TTCCAGATATATAACATAACATCCGCTCTGTAGACTTCAAAATACGGGACCTGCAGAG 1709

QY 172 GTTGGCATTGGAATACAG 191
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1710 ACAGGTACTGTGACTTACAG 1729

RESULT 10
LOCUS      AF052746      3471 bp      mRNA      ROD      08-APR-1999
DEFINITION Mus musculus chloride channel CACC mRNA, complete cds.
ACCESSION  AF052746
NID         93560546
VERSION     AF052746.1 GI:3560546
KEYWORDS
SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS     Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE       1 (bases 1 to 3471)
JOURNAL     Romolo, L., Musante, L., Cinti, R., Seri, M., Moran, O., Zegarra-Moran, O.
MEDLINE     and Galietta, L.J.V.
REFERENCE   2 (bases 1 to 3471)
AUTHORS     Romolo, L., Musante, L., Cinti, R., Seri, M., Moran, O., Zegarra-Moran, O.
TITLE       Characterization of a murine gene homologous to the bovine CACC
JOURNAL     chloride channel
MEDLINE     Gene 228 (1-2), 181-188 (1999)
REFERENCE   9913882
AUTHORS     Galietta, L.J.V.
TITLE       Direct Submision
JOURNAL     Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto
MEDLINE     Giannina Gaslini, Genova I-16148, Italy
REFERENCE   Location/Qualifiers
AUTHORS     1.3471
TITLE       /organism="Mus musculus"
JOURNAL     /db_xref="taxon:10090"
MEDLINE     /db_xref="dbEST:AA245677"
REFERENCE   47.2752
AUTHORS     /codon_start=1
TITLE       /product="chloride channel CACC"
JOURNAL     /protein_id="AAC35003.1"
MEDLINE     /db_xref="PID:g3560547"
```

```
/db_xref="GI:3560547"
/translation="MVGQLVLLFLTLHLQNTSSMHLNSGIVGVIAINPSVE
DELRILIPSEMYQASTYLEFAQGVYFRNLSILPVMKSRVILGMPRESYDAD
VIVADHLHGDDEPVLQYGGCGRQYHFTNPNFLTLNLRYPGRGRVFEAMHL
RMGFEVNDOPFYMRSKNTIETRCSTRITGTGNVYHNCERNGCYTRARBRDSTRL
YERKCTFIPIBKITOTAGASINIFEMONINSYVEFCETREKHNNAAPILQNMCRBRTWYI
KISADPOMAPMRKTEAPPPPTSLKSRRVCLVDKSGMDKEDRLRMQOAEVL
YLQIVKESMWGLVTFDSNAHIONLITITSSDYOKITANLPQASGGSICHGLQ
AGFOAITSDDOSTSGSEIYLLDGEDNGIRSCFEAVRSQAIHTITAGSARELET
LSMTGGLRFEYANKDLNSLIDAFSRISSTGSYSOQALQESKAFVBARAWINGTVP
LSDTVGNDFFEVITMVKRPEIILDPKCKKXTTSPDOKLNRASARLOIPGTAEFG
TWYITSTGTSQILMTVTTTRASPMEPLATAHMSQSTAOQPSRMIYARYSQOGL
PVLGAVNTALIEAHEGHQVTLMDNGAGADTVKNDGITYRTFDTHGNGRYSLSKRV
OQAKNTRLSLRQKNKSLYIPGVENGKIVLNPDPVOEALAEVDEDRYTSKGS
FTVSGAPPDGDHARVPPSKVLDLEAFEGIDYIHLTWAPKYLMDNGRAHRYIIRMSQ
HPDLQEDENNATLVNASSLIPREAGSKETFEKPEFETRIANGIOIYIAIQADNEASTL
TSEVSNIAQVKTLSLEDSISALGDISAISMTIMGLTIFNSILN"

BASE COUNT      1096 a      777 c      736 g      862 t
ORIGIN

Query Match      23.4%; Score 53.6; DB 12; Length 3471;
Best Local Similarity 58.5%; Pred. No. 5.2e-08;
Matches 117; Conservative 0; Mismatches 74; Indels 9; Gaps 1;

QY 1 GGCACAGTGAATCGTGACAGCACCCTGGGAAAGACACTTGTTCCTATCACCCTGGACA 60
    || ||||| ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 1559 GGTACAGTACTCTGACAGACGCTGGCAAGACAGCTTCTTGTATCACCCTGGATG 1618

QY 61 ACCGAGCCTCCCAATCTCTCTGGGATCCAGTGACAGAAAGCA-----AGGT 111
    || ||||| ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 1559 GTAAAAAAGCCAGAAATCATCTTCAAGATCCAAAGAAAATAATATACACCTCAGAT 1618

QY 112 GGGTTTGTAGTGACAAAACACCAAAATGGCCTACCTCCAAATCCAGGATGCTAAG 171
    || ||||| ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 1619 GTAAAAAAGCCAGAAATCATCTTCAAGATCCAAAGAAAATAATATACACCTCAGAT 1678

QY 172 GTTGGCATTGGAATACAG 191
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1739 ACAGGTACTGTGACTTACAG 1758

RESULT 11
LOCUS      AF043976      3415 bp      mRNA      PRI      07-APR-1999
DEFINITION Homo sapiens CLCA3 homolog (hCLCA3) mRNA, complete cds.
ACCESSION  AF043976
NID         94572288
VERSION     AF043976.1 GI:4572288
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens.
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS     Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE       1 (bases 1 to 3415)
JOURNAL     Gruber, A.D., and Pauli, B.U.
MEDLINE     Molecular cloning and biochemical characterization of a truncated,
REFERENCE   2 (bases 1 to 3415)
AUTHORS     Gruber, A.D., Eble, R.C. and Pauli, B.U.
TITLE       secreted member of the human family of Ca2+-activated Cl- channels
JOURNAL     Blochim. Biophys. Acta 1444 (3), 418-423 (1999)
MEDLINE     99196715
AUTHORS     Direct Submision
JOURNAL     Submitted (21-JAN-1998) Department of Pathology, College of
MEDLINE     Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
REFERENCE   Location/Qualifiers
AUTHORS     1.3415
TITLE       /organism="Homo sapiens"
JOURNAL     /db_xref="taxon:9606"
MEDLINE     /db_xref="hCLCA3"
REFERENCE   19.807
AUTHORS     /gene="hCLCA3"
TITLE       /gene="hCLCA3"
```

BASE COUNT	ORIGIN
1149 a	625 c 630 g 1011 t

Query Match	23.0%	Score 52.6	DB 42	Length 3415
Best Local Similarity	57.5%	Pred. No. 1.2e-07		
Matches 119	Conservative	0	Mismatches 79	Indels 9
				Gaps 1
Oy	1	GGCACGATATGCTGGACAGCACCGTGGGAAGACACTTGTTCCTATACCTGGACA	60	
Db	1588	GGTACAGCTCCCTGGATAGTACAGTTAAAGATACCTTCCTTGTGTGCATGGACG	1647	
Oy	61	ACGCAGCCTCCCAATTCCTCTCTGGGATCCCACTGACAGAAAGCAAGTGGCTTGT	120	
Db	1648	ATACAAAMCCAGCATAATTCCTCAAGATCCAAAGGAAAAATTTACTACTCGAT	1707	
Oy	120	-----AGTGACAAACACCAAAATGCGCTTACCTCCAAATCCAGCATGCTAAG	171	
Db	1708	TTTCAAGAAAGGTGACTAAATATTCGGTCTGCCCCGCTTGTGAATACCAAGTATTCGACAG	1767	
Oy	172	GTTGCACCTTGGAAATACAGTCTGAA	198	
Db	1768	ACAGGCACTTGGACTTACAGCGTTGA	1794	
RESULT 12				
LOCUS	AC005674			
DEFINITION	Homo sapiens chromosome 4, WORKING DRAFT SEQUENCE, 9 unordered pieces.			
ACCESSION	AC005674			
NID	94803911			
VERSION	AC005674.7			
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 220000)			
TITLE	Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.			
JOURNAL	Direct Submission			
REFERENCE	2 (bases 1 to 220000)			
AUTHORS	Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.			
TITLE	Direct Submission			
JOURNAL	Submitted (11-SEP-1998) Department of Genetics, Stanford Human Genome Center, 835 California Avenue, Palo Alto, CA 94304, USA			
COMMENT	On May 12, 1999 this sequence version replaced gi:4757680. * NOTE: This is a 'working draft' sequence. It currently * consists of 9 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.			
	1	1596: contig of 1596 bp in length		
	1597	2506: gap of unknown length		
	2507	4019: contig of 1513 bp in length		
	4020	4929: gap of unknown length		
	4930	6926: contig of 1997 bp in length		
	6927	7836: gap of unknown length		
	7837	9039: contig of 1203 bp in length		

	* 9040	9949:	gap of unknown length	
	* 9950	13819:	config of 3870 bp in length	
	* 13820	14728:	gap of unknown length	
	* 14730	18450:	config of 3721 bp in length	
	* 18451	19360:	gap of unknown length	
	* 19361	22742:	config of 3382 bp in length	
	* 22743	23651:	gap of unknown length	
*	23652	118804:	config of 95153 bp in length	
*	118805	119713:	gap of unknown length	
*	119714	220000:	config of 100287 bp in length.	
FEATURES	Location/Qualifiers			
source	1..220000			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="4"			
BASE COUNT	57281 a	49383 c	48355 g	57702 t 7279 others
ORIGIN				
Query Match		15.3%;	Score 35;	DB 35; Length 220000;
Best Local Similarity		49.2%;	Pred. No. 0.26;	
Matches	92; Conservative	0;	Mismatches	95; Indels 0; Gaps
OY 1	GGCACATGATCGTGGAGCACCCTGGGAAGCACACTTTGTTCATTACCTTGACA	60		
Dbl 193500	GGCAACATAAAGGGGCCCATCTCTACAAAAAACACAAAAGGTTCTGTGGCTCTTTTA	193559		
OY	61 ACGCAGCCCTCCCCAAATCTCTCTCTGGGATCCCCGTGAGACAGAAGAGSTGGCTTTGTA	120		
Dbl 193560	GGGCTCCCTCCAATTCACCTTTTTGTGAGATCCCATATACCTTTTGTGATGAGTAGAACTTA	193619		
OY	121 GTGGACAAAAACACCAAATATGGCCTTACCTCCAAATCCAGGCAATGCTAAGGTTGGCACT	180		
Dbl 193620	GTTTCACAAATGATGAAGTGTTCTCTCCTCTTAACACAGAAGTTTCTCACGCTTGGCACC	193679		
OY	181 TGGAAAT	187		
Dbl 193680	ATGACAT	193686		

```

RESULT 13
AB026833 LOCUS AB026833 3604 bp mRNA PRI 26-MAY-1999
DEFINITION Homo sapiens mRNA for chloride channel protein, complete cds.
ACCESSION AB026833
MID 94887600
VERSION AB026833.1 GI:4887600
KEYWORDS chloride channel protein.
SOURCE Homo sapiens adult corneal epithelium cDNA to mRNA,
clone_1lb:lambda ZAP clone:lambda 13802.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Itoh,R., Kawamoto,S., Kinoshita,S., Kawasaki,S., and Okubo,K.
TITLE Isolation and characterization of chloride channel in human corneal
epithelium
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 3604)
AUTHORS Itoh,R., Kawamoto,S., and Okubo,K.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1999) to the DDBJ/EMBL/Genbank databases. Reiko
Itoh, Institute for Molecular and Cellular Biology, Osaka
University, Yamada-oka 1-3, Suita, Osaka 565-0871, Japan
(E-mail:reiko@imcb.osaka-u.ac.jp, tel:81-6-6879-7992,
Fax:81-6-6877-1922)
FEATURES
Source
location/Qualifiers
1..3604
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="lambda 13802"
/clone_1lb="lambda ZAP"

```



CDS  
/dev\_stage="adult"  
/map="lp2"  
/tissue\_type="corneal epithelium"  
42..2873  
/codon\_start=1  
/product="chloride channel protein"  
/protein\_id="BA077810.1"  
/db\_xref="PDB:1O41578"  
/db\_xref="PIR:G4887601"  
/db\_xref="GI:4887601"  
/translation="MORSIAGPICNLKFTLLVALSSEPLIGAVOLQDNGYGL  
IAINPOVENOMLISNIKEMITEASFLYENAKRRVFEFNITLLIPATRKANNKIK  
QESTERKANYIVDMRGARHDDPYTIOYRCCKEGKTYHTPFLNDNTLAGSRGR  
VFEHMAHLRMVDEYINDKPPYINGQOIVTRCSSDPTIFVCEKPCQENCII  
SKLFEGCTFIYNSTONATASIMFQSLSSVEFCNASIHNEAPLQMCRLSASW  
DYITDSADFHHSFPMNGTELPPTFSLVQADKVCYLIDVSSKMAEDRLQLOQA  
AEFYLMQIVETHTFVGIAFSFGKEIRAOHQINSNDKRLVSYLPTVSAKTDISI  
CSGLKKEFEVVKLNGKAYGSVMILVTSGDRLGNCPLTVLSGSTIHSIALGSSAA  
PRLERLSRLTGKLFVDPDISNSNMIDFSRISSTGDIIPQOHIOLESTGENVPH  
QKNTVVDNTVGNDDTMEFLVWQSGPPEIILDPDGKRYTNFTITNLTFRSLMI  
PQTAPGHMTYTLNTHSLQALKVTVSRASNSAVPAPVAFVERDSLHPHPMI  
YANVRQGFYPIINATVATVETEDPTLRLDAGADVIKNDGISRYFFSFAAN  
GRYSLKVHNHSPISIPASISIPGSHAMVPGVTANGNIONNAPKRSVGRNEERKRG  
FSRVSSGGSFVLGVAPGHPDPKCKIIDEAVKVEEELTSMWAPGEDPDGOAT  
SEIEMSKSLQNIQDNNAILVNTSKRPPQAGIREITFSPQISTNGPEHONGET  
HESHRIYVAIRAMDNSLOSASVSNIAQAPLFTPPNSDVPARDYLLKGLVTAMGLIG  
IICLIIVVHTLSRRKRRADKNGTKTL"

BASE COUNT 1113 a 740 c 732 g 1017 t 2 others  
ORIGIN

Query Match 14.9%; Score 34.2; DB 9; Length 3604;  
Best Local Similarity 76.4%; Pred. No. 0.25;  
Matches 42; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 3 CACAGTATCGTGGACAGACCGTGGGAAAGACACTTGTCTTATCAGCTGG 57  
DB 1577 CACAGTACTGTGATTAATCTGTGGACAGACACTTGTCTTATCAGTGG 1631

RESULT 14  
CEY38E10.1/c  
WPCOMMENT  
Sequence split into 4 fragments LOCUS CEY38E10 Accession AL021149

Fragment Name Begin End  
CEY38E10\_0 1 110000  
CEY38E10\_1 100001 210000  
CEY38E10\_2 200001 310000  
CEY38E10\_3 300001 393707  
Continuation (2 of 4) of CEY38E10 from base 100001 (AL021149 Caenorhabditis elegans chrC)

Query Match 14.7%; Score 33.6; DB 34; Length 110000;  
Best Local Similarity 55.0%; Pred. No. 0.69;  
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 20 GCACCGTGGGAAAGACACTTGTCTTATCAGCTGGACAGCAGCCCTCCCAATCC 79  
DB 64430 GCACCTTGGGAAATTTTGGCGCCCTTTTCCGATTCATGATTAATGACTAGAC 64371  
OY 80 TTCTCTGGGATCCCATGAGACAGAGAGGTGCTTTGTAGTGACAAAAACACCAAAA 139  
DB 64370 TTCCCGGAGCCCAATCGATAAAGTAGATGCCAGGTGAGCGAGAAAAATACAAAA 64311

RESULT 15  
CEY57A10.4/c  
WPCOMMENT  
Sequence split into 5 fragments LOCUS CEY57A10 Accession AL020986  
Fragment Name Begin End  
CEY57A10\_0 1 110000  
CEY57A10\_1 100001 210000  
CEY57A10\_2 200001 310000

CEY57A10\_3 300001 410000  
CEY57A10\_4 400001 469134  
Continuation (5 of 5) of CEY57A10 from base 400001 (AL020986 Caenorhabditis elegans c

Query Match 14.7%; Score 33.6; DB 34; Length 69134;  
Best Local Similarity 55.0%; Pred. No. 0.64;  
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 20 GCACCGTGGGAAAGACACTTGTCTTATCAGCTGGACAGCAGCCCTCCCAATCC 79  
DB 20097 GCACCTTGGGAAATTTTGGCGCCCTTTTCCGATTCATGATTAATGACTAGAC 20038  
OY 80 TTCTCTGGGATCCCATGAGACAGAGAGGTGCTTTGTAGTGACAAAAACACCAAAA 139  
DB 20037 TTCCCGGAGCCCAATCGATAAAGTAGATGCCAGGTGAGCGAGAAAAATACAAAA 19978

Search completed: August 5, 1999, 23:13:06  
Job time: 7669 sec

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:44 ; Search time 3059.95 Seconds  
(without alignments)  
147.620 Million cell updates/sec

Title: US-09-049-696-10

Perfect score: 229  
Sequence: 1 GGCACTGATGCTGGACAG.....AACCTTGACCTGACTGTCA 229

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:\*

1: em\_est1:.\*  
2: em\_est2:.\*  
3: em\_est3:.\*  
4: em\_est4:.\*  
5: em\_est5:.\*  
6: em\_est6:.\*  
7: em\_est7:.\*  
8: em\_est8:.\*  
9: em\_est9:.\*  
10: em\_est10:.\*  
11: em\_est11:.\*  
12: em\_est12:.\*  
13: em\_est13:.\*  
14: em\_est14:.\*  
15: em\_est15:.\*  
16: em\_est16:.\*  
17: em\_est17:.\*  
18: em\_est18:.\*  
19: em\_est19:.\*  
20: em\_est20:.\*  
21: em\_est21:.\*  
22: em\_est22:.\*  
23: em\_est23:.\*  
24: em\_est24:.\*  
25: em\_est25:.\*  
26: em\_est26:.\*  
27: em\_est27:.\*  
28: em\_est28:.\*  
29: em\_est29:.\*  
30: em\_est30:.\*  
31: em\_est31:.\*  
32: em\_est32:.\*  
33: em\_est33:.\*  
34: em\_est34:.\*  
35: em\_est35:.\*  
36: em\_est36:.\*  
37: em\_est37:.\*  
38: em\_est38:.\*  
39: em\_est39:.\*  
40: em\_est40:.\*  
41: em\_est41:.\*  
42: em\_est42:.\*  
43: em\_est43:.\*  
44: em\_est44:.\*  
45: em\_est45:.\*  
46: em\_est46:.\*  
47: em\_est47:.\*  
48: em\_est48:.\*  
49: em\_est49:.\*  
50: em\_est50:.\*  
51: em\_est51:.\*  
52: em\_est52:.\*  
53: em\_est53:.\*

54: em\_est22:.\*  
55: em\_est23:.\*  
56: em\_est24:.\*  
57: em\_est25:.\*  
58: em\_est26:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	26	11.4	455	39	AA871197	AA871197 vq32c11.r
2	19	8.3	397	25	W16532	W16532 zb10b12.r1
3	19	8.3	293	31	AA321170	AA321170 ESTR23647
4	18	7.9	520	27	AA051927	AA051927 SMRFCA554
5	18	7.9	511	31	AA311390	AA311390 EST182307
6	18	7.9	316	34	AA486171	AA486171 ab14f02.r
7	18	7.9	404	36	AA659068	AA659068 nub1d02.s
8	17	7.4	416	20	D21547	D21547 M0568H09.mo
9	17	7.4	415	20	T05563	T05563 EST03452.Fe
10	17	7.4	417	20	T35174	T35174 EST81287.Hu
11	17	7.4	101	21	T61187	T61187 YC45C09.r1
12	17	7.4	543	21	T63273	T63273 YC08d10.s1
13	17	7.4	213	22	R79129	R79129 y186c12.s1
14	17	7.4	327	25	N68142	N68142 y255d11.s1
15	17	7.4	438	25	N68218	N68218 y256d11.s1
16	17	7.4	502	25	N76112	N76112 z486c05.s1
17	17	7.4	166	26	W18696	W18696 mcl10a04.r1
18	17	7.4	284	26	W70958	W70958 me24a02.r1
19	17	7.4	589	27	AA003372	AA003372 mg47909.r
20	17	7.4	629	29	AA166751	AA166751 z085c07.s
21	17	7.4	453	29	AA177303	AA177303 mt22f11.r
22	17	7.4	199	30	AA251352	AA251352 z508c02.s
23	17	7.4	337	30	AA262477	AA262477 z517908.s
24	17	7.4	275	32	AA344252	AA344252 EST50133
25	17	7.4	428	33	AA402294	AA402294 z446d06.r
26	17	7.4	385	33	AA429793	AA429793 z574d06.r
27	17	7.4	200	33	AA450328	AA450328 z804908.s
28	17	7.4	565	34	AA464650	AA464650 z85c03.s
29	17	7.4	317	34	AA464694	AA464694 z82a08.s
30	17	7.4	303	34	AA522658	AA522658 n17le07.s
31	17	7.4	644	34	AA531016	AA531016 n107c11.s
32	17	7.4	511	35	AA543238	AA543238 vK37910.r
33	17	7.4	385	35	AA597584	AA597584 29492.Lam
34	17	7.4	376	36	AA602425	AA602425 n096c10.s
35	17	7.4	326	36	AA639827	AA639827 ng84h06.s
36	17	7.4	591	36	AA662424	AA662424 nu97907.s
37	17	7.4	386	37	AA705660	AA705660 z4f0d12.s
38	17	7.4	565	38	AA775457	AA775457 z429a01.s
39	17	7.4	442	38	AA775467	AA775467 z429b01.s
40	17	7.4	376	39	AA831491	AA831491 oc61c01.s
41	17	7.4	313	39	AA834992	AA834992 a150c04.s
42	17	7.4	494	39	AA836071	AA836071 oc99f06.s
43	17	7.4	327	40	AA917401	AA917401 o180c07.s
44	17	7.4	304	41	AA995637	AA995637 os02f05.s
45	17	7.4	534	50	AI694335	AI694335 wd45f02.x

ALIGNMENTS

RESULT 1  
AA871197 455 bp mRNA EST 16-MAR-1998  
LOCUS vq32c11.r1 Barstead bowel MFLRB9 Mus musculus CDNA clone  
DEFINITION IMAGE:1095956 5' similar to SW:EDIC BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.  
ACCESSION AA871197  
NTD 92366642

```

VERSION      AA871197.1  GI:2966642
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus.
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 455)
              Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
              Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
              Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
              Theisinger,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and
              Watson,R.
TITLE        The WashU-HHMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1395394.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
GSI:602188
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 414.
Location/Qualifiers
1..455
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1095956"
/clone_1ib="Barstead bowel MPLRB9"
/tissue.type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: p7T73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACGACATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGAGCTCTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified p7T73 vector.
Source irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob
Barstead."

BASE COUNT   122 a      123 c      105 g      105 t
ORIGIN
Query Match      11.4%; Score 26; DB 39; Length 455;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      40 TTGTTCTTATCACCTGGACACGCA 65
|||||
Db      161 TTGTTCTTATCACCTGGACACGCA 186

RESULT      2
LOCUS       W16532
DEFINITION  w16532 397 bp mRNA EST 29-APR-1996
IMAGE:301631 5', mRNA sequence.
ACCESSION   W16532
NID         g1290914
VERSION     W16532.1  GI:1290914
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

REFERENCE
1 (bases 1 to 397)
AUTHORS
Eutheria; Primates; Catarrhini; Homnidae; Homo.
Hillier,L., Clark,N., Dubouque,T., Elliston,K., Hawkins,M.,
Holman,M., Hallman,M., Kueba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaslis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
TITLE
The WashU-Merck EST Project
JOURNAL
Unpublished (1995)
COMMENT
On May 8, 1995 this sequence version replaced gi:800920.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: EPrimer
High quality sequence stop: 317.

FEATURES
Source
Location/Qualifiers
1..397
/organism="Homo sapiens"
/db_xref="GDB:1246555"
/db_xref="taxon:9606"
/clone="IMAGE:301631"
/clone_lib="Soares fetal_lung_NbHL19w"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Lung; Vector: pRT3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-GTGTACCAATCTGACGAGCGAGCGCCGACATTTTTTTTTTTT-3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19w."
BASE COUNT
109 a 95 c 110 g 82 t 1 others
ORIGIN

Query Match 8.3%; Score 19; DB 25; Length 397;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 75 AATCCTCTCTCGGATCCC 93
|||||
Db 386 AATCCTCTCTCGGATCCC 368

RESULT 3
LOCUS AA321170 293 bp mRNA EST 19-APR-1997
DEFINITION EST223647 Frontal lobe Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA321170
NID g1973518
VERSION AA321170.1 GI:1973518
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 293)
ADAMS,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library
Nature Genet. 4, 373-380 (1993)
JOURNAL
MEDLINE
94004965

```

COMMENT On Sep 12, 1996 this sequence version replaced gi:1398058.

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M3 Reverse.

# FEATURES

source  
1..293  
/organism="Homo sapiens"  
/db\_xref="ATCC (lnhost):121740"  
/db\_xref="taxon:9606"  
/clone\_1lb="Frontal lobe"  
/dev\_stage="adult"  
/note="Organ: brain; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"  
BASE COUNT 103 a 42 c 88 g 58 t 2 others  
ORIGIN

Query Match 8.3%; Score 19; DB 31; Length 293;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 CAGCCTCCCAATCTTC 82  
|||||  
Db 154 CAGCCTCCCAATCTTC 136

RESULT 4  
AA051927 520 bp MRNA EST 12-DEC-1996  
LOCUS SWMFCAS54SR Brugia malayi microfilaria cDNA (SWM94LS-Bmmf) Brugia  
DEFINITION malayi cDNA clone SWMFCAS54 5', mRNA sequence.  
ACCESSION AA051927  
NID 91537094  
VERSION AA051927.1 GI:1537094  
KEYWORDS EST.  
SOURCE Brugia malayi.  
ORGANISM Brugia malayi.  
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;  
Filarioidae; Onchocercidae; Brugia.  
REFERENCE 1 (bases 1 to 520)  
AUTHORS Williams, S.A.  
TITLE Genes expressed in microfilaria of Brugia malayi  
JOURNAL Unpublished (1995)  
COMMENT

Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genomesmith.edu  
Seq primer: pBluescript SK.  
Location/Qualifiers  
1..520  
/organism="Brugia malayi"  
/strain="TIGR Labs"  
/db\_xref="taxon:6279"  
/clone="SWMFCAS54"  
/clone\_1lb="Brugia malayi microfilaria cDNA  
(SWM94LS-Bmmf)"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: lambda Unizap XR; Site\_1: EcoR I; Site\_2:  
Xho I; Lymphatic filarial nematode parasite of humans.

# FEATURES

BASE COUNT 146 a 59 c 85 g 221 t 9 others  
ORIGIN

Query Match 7.9%; Score 18; DB 27; Length 520;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 TTGGCACTGGGAATACA 190  
|||||  
Db 373 TTGGCACTGGGAATACA 390

RESULT 5  
AA11390 511 bp MRNA EST 19-APR-1997  
LOCUS EST182307 Jurkat T-cells VI Homo sapiens cDNA 5' end, mRNA  
DEFINITION sequence.  
ACCESSION AA11390  
NID 91963904  
VERSION AA11390.1 GI:1963904  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 511)  
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Feldner, R.A.,  
Bult, C.J., Lee, N.H., Kinkness, E.F., Weissstock, K.G., Gocayne, J.D.,  
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, W.A., C.,  
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,  
Fitzgerald, L.M., Fitzhugh, W.M., Fitchman, J.L., Geoghegan, N.S.,  
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,  
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
Moreno-Palanges, R.F., McDonald, L.A., Nguyen, D.T., Pellierino, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,  
Small, R.V., Spriggs, T.A., Utecht, T.R., Weidman, J.F., Li, Y.,  
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,  
He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,  
Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H.,  
Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G., Ruben, S.M.,  
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,  
Fraser, C.M., and Venter, J.C.  
TITLE Initial assessment of human gene diversity and expression patterns  
JOURNAL based upon 83 million nucleotides of cDNA sequence  
MEDLINE Nature 377 (6547 Suppl.), 3-174 (1995)  
COMMENT

On Sep 12, 1996 this sequence version replaced gi:1400909.  
Other ESTs: TRC184526  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M3 Reverse.

Location/Qualifiers  
1..511  
/organism="Homo sapiens"  
/db\_xref="ATCC (lnhost):158802"

# FEATURES

source

```
/db_xref="taxon:9606"
/clone_lib="Jurkat T-cells VI"
/cell_type="T-lymphocyte"
/note="Vector: pBluescript SK-; site_1: EcoRI; site_2:
XhoI"
BASE COUNT      124 a      136 c      157 g      92 t      2 others
ORIGIN

Query Match      7.9%; Score 18; DB 31; Length 511;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 ATCCTTCTGTGGATCCC 93
|||||
Db 410 ATCCTTCTGTGGATCCC 393

RESULT 6
AA486171/c 316 bp mRNA EST 06-MAR-1998
LOCUS ab14f02.r1 Stratogene lung (#937210) Homo sapiens cDNA clone
DEFINITION IMAGE:840795 5', mRNA sequence.
ACCESSION AA486171
NID 92216387
VERSION AA486171.1 GI:2216387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 316)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Thaiesing,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
Mashu-NCI human EST Project
Unpublished (1997)
JOURNAL On Apr 14, 1993 this sequence version replaced gi:692847.
COMMENT

FEATURES
source
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 889 Std Error: 0.00
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 200.
Location/Qualifiers
1. 316
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="814H06: 14"
/clone="IMAGE:840795"
/clone_lib="Stratogene lung (#937210)"
/sex="male"
/dev_stage="72 years"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; site_1:
EcoRI; site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT, normal lung. Average insert size: 1.0 kb;
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCGCGAG
3' -3' adaptor sequence: 5' CTCGAGCTTTT TTTT TTTT TTTT 3'"
BASE COUNT      84 a      81 c      103 g      48 t
ORIGIN

Query Match      7.9%; Score 18; DB 34; Length 316;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 76 ATCCTTCTGTGGATCCC 93
|||||
Db 150 ATCCTTCTGTGGATCCC 133

RESULT 7
AA659068 404 bp mRNA EST 03-DEC-1997
LOCUS nu81d02.s1 NCI_CGAP_AlV1 Homo sapiens cDNA clone IMAGE:1217091,
DEFINITION mRNA sequence.
ACCESSION AA659068
NID 92595222
VERSION AA659068.1 GI:2595222
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL On May 5, 1995 this sequence version replaced gi:798066.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bdrrp/image/image.html
Insert Length: 949 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 396.
Location/Qualifiers
1. 404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone="IMAGE:1217091"
/clone_lib="NCI_CGAP_AlV1"
/tissue_type="alveolar rhabdomyosarcoma"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from alveolar
rhabdomyosarcoma, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Reference: Krizman et al.
(1996) Cancer Research 56:5380-5383."
BASE COUNT      104 a      104 c      140 g      56 t
ORIGIN

Query Match      7.9%; Score 18; DB 36; Length 404;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 ATCCTTCTGTGGATCCC 93
|||||
Db 154 ATCCTTCTGTGGATCCC 137

RESULT 8
D21547/c 416 bp mRNA EST 07-OCT-1996
LOCUS M0568H09 mouse embryonal carcinoma cell line F9 Mus musculus cDNA
DEFINITION clone 68H09, mRNA sequence.
```

ACCESSION D21547  
 NID 9618684  
 VERSION D21547.1 GI:618684  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 416)  
 AUTHORS Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T. and Shimada, K.  
 TITLE A survey of genes expressed in undifferentiated mouse embryonal carcinoma F9 cells: characterization of low-abundance mRNAs  
 JOURNAL J. Biochem. 116, 128-139 (1994)  
 MEDLINE 95096008  
 COMMENT Contact: Kazunori Shimada  
 Department of Medical Genetics, Division of Molecular Biomedicine  
 Research Institute for Microbial Diseases, Osaka University  
 3-1, Yamadaoka, Suita, Osaka, 565, Japan  
 Tel: 06-879-8325  
 Fax: 06-879-8326.

FEATURES  
 source  
 1..416  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_11b="mouse embryonal carcinoma cell line F9"

BASE COUNT 87 a 96 c 116 g 117 t  
 ORIGIN

Query Match 7.4%; Score 17; DB 20; Length 416;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 CCTCTCTGGATCCCA 94  
 ||||||||||||||||  
 Db 353 CCTCTCTGGATCCCA 337

RESULT 9  
 T05563 415 bp mRNA EST 30-JUN-1993  
 LOCUS EST03452 Fetal brain, Striatogene (cat#936206) Homo sapiens cDNA  
 DEFINITION clone HFBDC39, mRNA sequence.  
 ACCESSION T05563  
 NID G316713  
 VERSION T05563.1 GI:316713  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 415)  
 AUTHORS Adams, M.D., Kerlavage, A.R., Fields, C. and Venter, J.C.  
 TITLE 3,400 expressed sequence tags identify diversity of transcripts from human brain  
 JOURNAL Nature Genet. 4, 256-267 (1993)  
 MEDLINE 93364420  
 COMMENT Contact: Adams, MD  
 The Institute for Genomic Research  
 932 Clopper Road, Gaithersburg, MD 20878  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: mdadams@tigr.org  
 Seq primer: M13-21.  
 Location/Qualifiers  
 1..415  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (lnhost):82263"  
 /db\_xref="taxon:9606"  
 /clone="HFBDC39"

BASE COUNT 139 a 64 c 79 g 128 t  
 ORIGIN

Query Match 7.4%; Score 17; DB 20; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 ACAACCTTGACCTCGA 223  
 ||||||||||||||||  
 Db 292 ACAACCTTGACCTCGA 308

RESULT 10  
 T35174 417 bp mRNA EST 06-SEP-1995  
 LOCUS EST81287 Human prostate gland Homo sapiens cDNA 3' end similar to  
 DEFINITION None, mRNA sequence.  
 ACCESSION T35174  
 NID G617272  
 VERSION T35174.1 GI:617272  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 417)  
 AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulmer, R.A., Bult, C.J., Lee, N., Kirness, E.F., Weinstein, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.-W., Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S.M., Glodex, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkley, P.S., Kelley, J.M., Kline, K.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merriell, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pellegrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.-J., Dimke, D., Feng, P., Ferrie, A., Fischer, C., Hastings, G.A., He, W.-W., Hu, J.-S., Greene, J.M., Gruber, J., Hudson, P., Kim, A., Kozak, D.L., Kunsch, C., Li, H., Melsner, P.S., Olsen, H., Raymond, L., Wei, Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.  
 TITLE Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million Basepairs of cDNA Sequence  
 JOURNAL Nature 377, 3-174 (1995)  
 MEDLINE 96026280  
 COMMENT Other ESTs: EST81288 THC10759  
 Contact: Venter, JC  
 The Institute for Genomic Research  
 932 Clopper Rd, Gaithersburg, MD 20878  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: tdbinfo@tigr.org  
 For clone availability, additional sequence and expression information related to this EST, please contact the TIGR Database (tdbinfo@tigr.org)  
 Seq primer: M13-21.  
 Location/Qualifiers  
 1..417  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (lnhost):106371"  
 /db\_xref="taxon:9606"  
 /clone\_11b="Human Prostate gland"  
 /note="Organ: prostate gland"

BASE COUNT 119 a 111 c 82 g 105 t  
 ORIGIN

Query Match 7.4%; Score 17; DB 20; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 150 CCAATCCAGGACATG 166  
 DB 10 CCAATCCAGGACATG 26

## RESULT 11

LOCUS T61187/c

DEFINITION T61187 101 bp mRNA EST 13-FEB-1995  
 y45c09.r1 Strata gene liver (#937224) Homo sapiens cDNA clone  
 IMAGE:83632 5' similar to contains Alu repetitive element; mRNA  
 sequence.

ACCESSION T61187  
 NID 9664224  
 VERSION T61187.1 GI:664224  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 101)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Chissoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
 and Warr, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags  
 JOURNAL Genome Res. 6 (9), 807-828 (1996)  
 MEDLINE 97044478  
 COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Insert Size: 70

High quality sequence stops: 83 Source: IMAGE Consortium, LNL This  
 clone is available royalty-free through LNL; contact the IMAGE  
 Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 70 Std Error: 0.00  
 Seq primer: M13RPI

High quality sequence stop: 83.

## FEATURES

Source

Location/Qualifiers  
 1..101  
 /organism="Homo sapiens"  
 /db\_xref="GDB:500689"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:83632"  
 /clone\_lib="Stratagene liver (#937224)"  
 /sex="male"  
 /dev\_stage="49 years old"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: liver; Vector: pBluescript SK-; Site: 1;  
 EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. Hepatectomy from normal male caucasian. Average  
 insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor  
 sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'  
 CTCGAGTTTTTTTTTTT 3'."

BASE COUNT 26 a 24 c 28 g 23 t

ORIGIN

Query Match 7.4%; Score 17; DB 21; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 CAGCCTCCCAATCCT 80  
 DB 40 CAGCCTCCCAATCCT 24

## RESULT 12

LOCUS T63273

DEFINITION T63273 543 bp mRNA EST 17-FEB-1995  
 y008d10.s1 Strata gene lung (#937210) Homo sapiens cDNA clone  
 IMAGE:80083 3' similar to contains Alu repetitive element; mRNA  
 sequence.

ACCESSION T63273  
 NID 9667138  
 VERSION T63273.1 GI:667138  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 543)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Chissoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
 and Warr, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags  
 JOURNAL Genome Res. 6 (9), 807-828 (1996)  
 MEDLINE 97044478  
 COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Insert Size: 592

High quality sequence stops: 424 Source: IMAGE Consortium, LNL This  
 clone is available royalty-free through LNL; contact the IMAGE  
 Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 592 Std Error: 0.00  
 Seq primer: -2im13

High quality sequence stop: 424.

## FEATURES

Source

Location/Qualifiers  
 1..543  
 /organism="Homo sapiens"  
 /db\_xref="GDB:483700"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:80083"  
 /clone\_lib="Stratagene lung (#937210)"  
 /sex="male"  
 /dev\_stage="72 years"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: lung; Vector: pBluescript SK-; Site: 1;  
 EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. normal lung. Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGAG  
 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'." 14 others

BASE COUNT 134 a 152 c 106 g 137 t

ORIGIN

Query Match 7.4%; Score 17; DB 21; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 146 ACCTCAATCCAGGC 162  
 DB 67 ACCTCAATCCAGGC 83

RESULT 13



R79129  
LOCUS 213 bp mRNA EST 09-JUN-1995  
DEFINITION y186c12.s1 Soares placenta Nb2HP Homo sapiens cDNA clone  
IMAGE:146134 3', mRNA sequence.  
ACCESSION R79129  
NID 9855410  
VERSION R79129.1 GI:855410  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 213)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On May 8, 1995 this sequence version replaced g1:799900.

TITLE  
JOURNAL  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
High quality sequence stops: 137  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: Promega -21ml3  
High quality sequence stop: 137.  
Location/Qualifiers  
1. 213  
/organism="Homo sapiens"  
/db\_xref="GDB:557642"  
/db\_xref="taxon:9606"  
/clone="IMAGE:146134"  
/clone.lib="Soares placenta Nb2HP"  
/sex="Female"  
/dev\_stage="placenta obtained at birth (full term)"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Placenta; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
AAGTCGAGAAATTCGCGCCGAGAAATTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 80 a 30 c 29 g 5 t  
ORIGIN

Query Match 7.4%; Score 17; DB 22; Length 213;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 124 GACAAAACACCAAAAT 140  
|||||  
Db 6 GACAAAACACCAAAAT 22

RESULT 14  
LOCUS N68142 327 bp mRNA EST 13-MAR-1996  
DEFINITION y256d11.s1 Morton Fetal Cochlea Homo sapiens cDNA clone  
IMAGE:286965 3', mRNA sequence.  
ACCESSION N68142  
NID g1224303

VERSION N68142.1 GI:1224303  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 327)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,  
Chissie, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT On Nov 29, 1993 this sequence version replaced g1:502211.

TITLE  
JOURNAL  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: ml3 -40 forward  
High quality sequence stop: 169.  
Location/Qualifiers  
1. 327  
/organism="Homo sapiens"  
/db\_xref="GDB:3894082"  
/db\_xref="taxon:9606"  
/clone="IMAGE:286965"  
/clone.lib="Morton Fetal Cochlea"  
/tissue\_type="cochlea"  
/dev\_stage="16-22 week fetus"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: ear; Vector: pBluescript SK-; Site 1: EcoRI;  
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned  
unidirectionally. Primer: 0190 dT. Fetal cochlea, normal.  
37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP  
XR Vector. Library constructed by N. Robertson, C. Morton.  
-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor  
sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' "

BASE COUNT 115 a 47 c 50 g 11 t 4 others  
ORIGIN

Query Match 7.4%; Score 17; DB 25; Length 327;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 ACAACCTTGACCTGA 223  
|||||  
Db 160 ACAACCTTGACCTGA 176

RESULT 15  
LOCUS N68218 438 bp mRNA EST 13-MAR-1996  
DEFINITION y256d11.s1 Morton Fetal Cochlea Homo sapiens cDNA clone  
IMAGE:287061 3', mRNA sequence.  
ACCESSION N68218  
NID g1224379  
VERSION N68218.1 GI:1224379  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 438)

## AUTHORS

Hallier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,  
 Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, M.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J.,  
 Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
 and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)

JOURNAL  
 MEDLINE  
 97044478

COMMENT On Apr 14, 1993 this sequence version replaced gi:504450.

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium. (info@image.lnl.gov) for further information.

Seq primer: ml3 -40 forward  
 High quality sequence stop: 384.

Location/Qualifiers

1. 438  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3894178"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:287061"  
 /clone\_lib="Morton Fetal Cochlea"  
 /tissue\_type="cochlea"  
 /dev\_stage="16-22 week fetus"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: ear; Vector: pBluescript SK-; Site\_1: EcoRI;  
 Site\_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned  
 unidirectionally. Primer: 0190 dt. Fetal cochlea, normal.  
 37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP  
 XR Vector. Library constructed by N. Robertson, C. Morton.  
 -5' adaptor sequence: 5' GATTCGCGACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

BASE COUNT 153 a 67 c 62 g 155 t 1 others  
 ORIGIN

Query Match 7.4%; Score 17; DB 25; Length 438;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 ACAACCTTGACCTGA 223  
 ||||||||||||||||

DB 159 ACAACCTTGACCTGA 175

Search completed: August 6, 1999, 09:22:47  
 Job time: 6238 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:24:22 ; Search time 3264.21 Seconds

(without alignments)  
2906.327 Million cell updates/sec

Title: US-09-049-696-20

Perfect score: 2983  
Sequence: 1 GAAATCAGCGGAGACTGTAC.....AAATGCTAACACTGGGTA 2983

Scoring table:

Searched: 679419 seqs, 1590154680 residues

Database :

GeneBml:\*  
1: gb\_bal:\*  
2: gb\_bal:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_pat:\*  
6: gb\_pat:\*  
7: gb\_pat:\*  
8: gb\_pat:\*  
9: gb\_pat:\*  
10: gb\_pat:\*  
11: gb\_pat:\*  
12: gb\_pat:\*  
13: gb\_pat:\*  
14: gb\_pat:\*  
15: gb\_pat:\*  
16: gb\_pat:\*  
17: gb\_pat:\*  
18: gb\_pat:\*  
19: gb\_pat:\*  
20: gb\_pat:\*  
21: gb\_pat:\*  
22: gb\_pat:\*  
23: gb\_pat:\*  
24: gb\_pat:\*  
25: gb\_pat:\*  
26: gb\_pat:\*  
27: gb\_pat:\*  
28: gb\_pat:\*  
29: gb\_pat:\*  
30: gb\_pat:\*  
31: gb\_pat:\*  
32: gb\_pat:\*  
33: gb\_pat:\*  
34: gb\_pat:\*  
35: gb\_pat:\*  
36: gb\_pat:\*  
37: gb\_pat:\*  
38: gb\_pat:\*  
39: gb\_pat:\*  
40: gb\_pat:\*  
41: gb\_pat:\*  
42: gb\_pat:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2966.2	99.4	3311	11	AF039400	Homo sapi

2	2806.8	94.1	2826	42	AF127036	AF127036 Homo sapi
3	1764	59.1	2937	12	AB017156	AB017156 Mus muscu
4	933.4	31.3	2984	3	BF036445	U36445 Bos tauru
5	900.2	30.2	3317	3	AF001261	AF001261 Bos tauru
6	863	28.9	3022	12	AF047838	AF047838 Mus muscu
7	863	28.9	3471	12	AF052746	AF052746 Mus muscu
8	831	27.9	3415	42	AF043976	AF043976 Homo sapi
9	806	27.0	2820	3	AF001263	AF001263 Bos tauru
10	800.6	26.8	3288	3	AF001262	AF001262 Bos tauru
11	790.8	26.5	878	5	195746	195746 Sequence 8
12	594.4	19.9	35278	11	AF039401	AF039401 Homo sapi
13	550.6	18.5	3604	9	AB026833	AB026833 Homo sapi
14	381	12.8	1265	3	AF001264	AF001264 Bos tauru
15	32	7.218	5	166494	166494 Sequence 14	166494 Sequence 14
16	46.4	1.6	159453	34	AC006732	AC006732 Caenorhab
17	45.6	1.5	99083	36	CEY37A1B	AL023835 Caenorhab
18	43.8	1.5	256172	35	AC005139	AC005139 Plasmodiu
19	43.8	1.5	284972	35	AC005505	AC005505 Plasmodiu
20	43.2	1.4	113930	17	HSGEND	X64346 Herpesvyrus
21	43.2	1.4	43658	17	HSV3PERGEN	X64346 Herpesvyrus
22	43.2	1.4	110000	34	PFMAL4P1_0	M86409 Herpesvyrus
23	43.2	1.4	219200	35	AC006279	AL034557 Plasmodiu
24	43	1.4	91780	8	AC006192	AC006192 Plasmodiu
25	43	1.4	110000	34	PFMAL4P1_0	AL034557 Plasmodiu
26	42.4	1.4	164399	36	PFMAL3P6	298551 Plasmodiu
27	42.2	1.4	101679	8	ATF19B15	AL078470 Arabidops
28	42.2	1.4	97713	8	ATF25024	AL078469 Arabidops
29	42	1.4	126356	34	AC006876	AC006876 Caenorhab
30	42	1.4	258033	35	AC005507	AC005507 Plasmodiu
31	42	1.4	38174	37	CELE02H9	AF099915 Caenorhab
32	41.8	1.4	98067	9	HS435A7	AL022148 Homo sapi
33	41.8	1.4	216293	11	HTCRBCHR9	AF029308 Homo sapi
34	41.8	1.4	4601	36	DMU11584	U11584 Drosophila
35	41.8	1.4	19517	37	DMU37541	U37541 Drosophila
36	41.6	1.4	1231	1	MC082	Z33065 M. capricolu
37	41.6	1.4	119704	1	OSCHLPLXX	Z67753 O. siensis
38	41.6	1.4	228625	35	AC004153	AC004153 Plasmodiu
39	41.6	1.4	300172	35	AC005308	AC005308 Plasmodiu
40	41.6	1.4	220475	35	AC006278	AC006278 Plasmodiu
41	41.6	1.4	231699	35	AC006281	AC006281 Plasmodiu
42	41.4	1.4	53778	8	TIJ1	AF128393 Arabidops
43	41.4	1.4	154902	11	AC005799	AC005799 Homo sapi
44	41.4	1.4	199606	35	AC004688	AC004688 Plasmodiu
45	41.4	1.4	196149	35	AC004709	AC004709 Plasmodiu

#### ALIGNMENTS

RESULT	1	AF039400	3311 bp	mRNA	PRI	15-DEC-1998
LOCUS		AF039400	3311 bp	mRNA		15-DEC-1998
DEFINITION		Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,				
ACCESSION		AF039400				
KEYWORDS		complete cds.				
VERSION		94009457				
KEYWORDS		AF039400.1 GI:4009457				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS		Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE		Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.				
JOURNAL		Genomic cloning, molecular characterization, and functional				
MEDLINE		analysis of human CLCA1, the first human member of the family of				
REFERENCE		Ca2+-activated Cl- channel proteins				
AUTHORS		Genomics 54 (2), 200-214 (1998)				
TITLE		Gruber,A.D., Elble,R. and Pauli,B.U.				
		Direct Submission				

JOURNAL Submitted (19-Dec-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES

Location/Qualifiers  
1..3311

source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1..3311  
/gene="hCLCA1"  
352..3096  
/gene="hCLCA1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PIR:94009458"  
/db\_xref="GI:4009458"

CDS

```

//translation="MGPKSSVFLLHLLEGALSNSLIQLNNNGIYVAIDPNV
EDFTLIQIKDMVTQASLYLEATGKRFRFKNVALILPEWKTADYVRLPRLTKNA
DVLVAESTPGNDPEPYEQMGNGCEKBERHLTPDLAKKRLAEYQOGKAFVEMNAH
LRMGVPEYNDDEKFLSNGRIQAVRSAGITGNVVKCGSGCYTRKCFNKTGL
YEKGEVLOSROTERKASIMPAQVDSIVECTEONHNKEAPNOKNCRSWEYI
RSDSEFKTTPMTTOPPMPPSILQIGRIYCLVDKSGSATGNRLNRLNNGOFL
LOVEIGSWGMVTFDSAAYHOSLQIINSGSDPDLAKRLPAASGSGTSCSLRA
FTVIRKRYPTDGSSEIVLLTDEEDNTISGCRFEVYKOSGALITHVALGSAOELELSK
MTGKLOTASQVQVNNGLIDAFGALSSGNVRSIQLESKGLTLLNSOMMGTIVY
DSTVGKDTLELITWITOPQIILMDPSGQKOGGVQVKNRMALQIPJAKVGMVY
SLVASQTLTLTVSRASNALPITVTSKINKSPSPFLVYANIRGASILRA
SVTALIESVNGKIVTLELDNAGADATKDGVSRYFTYDNGRYSVRLAGVN
AARRVPIPOOSGALYIPGMLENDIOMNPPPEINRNDYOHKQVCSRTSGSFPAS
DVPRAPVLDLPPPGQIITDKRAIHGSLINLTWTPAPDDYDHGNAHYITIRISITLD
LRDFNSLQVNTTALIPKEANSEVLFRENTTFENGDLTALQAVKVDLKSFI
SNARVSLFIPQPTPEPSPDETSAACPNIHNSITPGIHLIKMKMWIGELQSLIA

```

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 99.4%; Score 2966.2; DB 11; Length 3311;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2979; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

QY 1 GAATTCACAGGAGATGTACAGCAATGGGCAATTAGAGTCTGTGTTGATT 60
DB 328 GGAATTCACAGGAGATGTACAGCAATGGGCAATTAGAGTCTGTGTTGATT 387
QY 61 CTTCACCTTCTAGAAAGGGCCCTAGTAATTCAGTCAATTCAGTGAACAACATGGCTAT 120
DB 388 CTTCACCTTCTAGAAAGGGCCCTAGTAATTCAGTCAATTCAGTGAACAACATGGCTAT 447
QY 121 GAAGGCATGTGCTGTCATGACGCCCAATGTCCAGAAAGATGAACAATCTATCAACAA 180
DB 448 GAAAGGCATGTGCTGTCATGACGCCCAATGTCCAGAAAGATGAACAATCTATCAACAA 507
QY 181 ATTAAGAGCATGTGACCCAGGACATCTGTATCTGTTGAAGTATAGGAAAGCGATT 240
DB 508 ATTAAGAGCATGTGACCCAGGACATCTGTATCTGTTGAAGTATAGGAAAGCGATT 567
QY 241 TATTTCAAAAAATGTTGCCATTTTATTCCTGAACAATGGAAGAAAGCGTATATGTG 300
DB 568 TATTTCAAAAAATGTTGCCATTTTATTCCTGAACAATGGAAGAAAGCGTATATGTG 627
QY 301 AGACCAAAACTGTGAGCACTACAAAAATGCTGATTTCTGTTGCTGAGTCTACTCTCCA 360
DB 628 AGACCAAAACTGTGAGCACTACAAAAATGCTGATTTCTGTTGCTGAGTCTACTCTCCA 687
QY 361 GGTAATGATGAACCTTACACTGAGCAGATGGGCAACTGTGGAGAGAAAGGGAAGATC 420
DB 688 GGTAATGATGAACCTTACACTGAGCAGATGGGCAACTGTGGAGAGAAAGGGAAGATC 747
QY 421 CACCTCACTGCTGATTTCAATTCAGAGAAAAAAGTTAGCTGAATATGACCAACAAGTAGG 480

```

```

DB 748 CACCTCACTGCTGATTTCAATTCAGAGAAAAAAGTTAGCTGAATATGACCAACAAGTAGG 807
QY 481 GCATTTCCTCATGATGGGCTCATCTACATGAGGAGATATTTGACGAGTCAATATGAT 540
DB 808 GCATTTCCTCATGATGGGCTCATCTACATGAGGAGATATTTGACGAGTCAATATGAT 867
QY 541 GAGAAATCTACTATTCATCAATGGAAGATACAAAGCAGTAAAGATGTTGAGGAGTATCT 600
DB 868 GAGAAATCTACTATTCATCAATGGAAGATACAAAGCAGTAAAGATGTTGAGGAGTATCT 927
QY 601 GGTCAAAATGATGTAAGAAGTGTACAGGAGCAGCTGTTACACCAAAAGATGCACATTC 660
DB 928 GGTCAAAATGATGTAAGAAGTGTACAGGAGCAGCTGTTACACCAAAAGATGCACATTC 987
QY 661 AATAAGTAAAGAGACCTCTATGAAAAAGATGTGAGTTTCTCCAAATCCCGCAGAGC 720
DB 988 AATAAGTAAAGAGACCTCTATGAAAAAGATGTGAGTTTCTCCAAATCCCGCAGAGC 1047
QY 721 GAGAAAGCTTCTATTAATGTTTGACACACATGTTGATTCATAGTTGAAATCTGTACGAA 780
DB 1048 GAGAAAGCTTCTATTAATGTTTGACACACATGTTGATTCATAGTTGAAATCTGTACGAA 1107
QY 781 CAATAACCAACAAAGAAAGAGCTCCAAACACAAATCAAAATGCAATCTCCAGACACA 840
DB 1108 CAATAACCAACAAAGAAAGAGCTCCAAACACAAATCAAAATGCAATCTCCAGACACA 1167
QY 841 TGGGAAGTGTATCGGTATGAGTGTGAGACTTTAAGAAACACCTCTATGACACACAGCA 900
DB 1168 TGGGAAGTGTATCGGTATGAGTGTGAGACTTTAAGAAACACCTCTATGACACACAGCA 1227
QY 901 CCAATATCCACTTCTTCATTTGCTGCAATGGAATGGAACAAAGATTTGTGTTAGTCTTGAC 960
DB 1228 CCAATATCCACTTCTTCATTTGCTGCAATGGAATGGAACAAAGATTTGTGTTAGTCTTGAC 1287
QY 961 AAATCTGAAGCAATGGGAGATGGTAAACCGCTCAATCGATGAATAAAGAGGCGAGCT 1020
DB 1288 AAATCTGAAGCAATGGGAGATGGTAAACCGCTCAATCGATGAATAAAGAGGCGAGCT 1347
QY 1021 TTCCTGTGAGACAGTGTGAGTGGGCTCTGGGTTGGGATGGTGAATTTGACATGCT 1080
DB 1348 TTCCTGTGAGACAGTGTGAGTGGGCTCTGGGTTGGGATGGTGAATTTGACATGCT 1407
QY 1081 GCCCATGTACAAAGTACATCATACAGATTAACAGTGGCAGTACAGGACACACTCGCC 1140
DB 1408 GCCCATGTACAAAGTACATCATACAGATTAACAGTGGCAGTACAGGACACACTCGCC 1467
QY 1141 AAAAGATTACCTGACAGCTTACAGAGGAGAGTCCATCTGACAGCGGCTTGGATGGCA 1200
DB 1468 AAAAGATTACCTGACAGCTTACAGAGGAGAGTCCATCTGACAGCGGCTTGGATGGCA 1527
QY 1201 TTTACTGTGATTGGAAGAAATATCCAACTGATGATCTGAAATGTGCTGTGAGGAT 1260
DB 1528 TTTACTGTGATTGGAAGAAATATCCAACTGATGATCTGAAATGTGCTGTGAGGAT 1587
QY 1261 GGGGAGAGCAACACTAATGAGTGGTCTTTAAGAGGTAAACAAAGTGTGCTCATATC 1320
DB 1588 GGGGAGAGCAACACTAATGAGTGGTCTTTAAGAGGTAAACAAAGTGTGCTCATATC 1647
QY 1321 CACACAGTGTGTTGGGGCTCTGAGCTCAAGAACTAGAGAGCTGTCCAAATAGACA 1380
DB 1648 CACACAGTGTGTTGGGGCTCTGAGCTCAAGAACTAGAGAGCTGTCCAAATAGACA 1707
QY 1381 GGAAGTTTACAGACATATGCTTACATCAAGTTCAGAACAAATGCGCTCATTTGATCTTT 1440
DB 1708 GGAAGTTTACAGACATATGCTTACATCAAGTTCAGAACAAATGCGCTCATTTGATCTTT 1767
QY 1441 GGGGCGCTTCAATCAAGAAATGAGTGTCTCAAGCGCTCCATCAGCTTGAAGTAAAG 1500
DB 1768 GGGGCGCTTCAATCAAGAAATGAGTGTCTCAAGCGCTCCATCAGCTTGAAGTAAAG 1827
QY 1501 GGATTAAACCTCAGAACAGCAGTGAATGATGACAGTATGCTGGACAGCACCGTG 1560

```



DSYKDTLFLITWTPOTLILMDPSOGKGFVNDKMTKATQIBKAKSTWXY  
SLQSSQTLITITVSRASNATLPITTYTSKNTKTSFPSPSLVYTAIRQASPLIRA  
SVAILBSVNGKRTIVLELLDNGADAGADKDVYSRFTTYDINGRSVRYALGYN  
AARRRVIPOOGALYIPQWIMENDEIOWNPBPPEINKDVOYKQVCFSTSGSFSVS  
DVPNAPIDPLEPQOIDLKAEIHGSLINLTWPAIGDDYHGHAKHIIIRISTIID  
LRDRENSLOVNTALIPKEANSEVEFEKRENTLEFEGTDLFAIQAVIDDKSEI  
SNIRVSLFIPQTPETPSPDETSAPOPNHINSTIPGHIILKIMKWIQELDLSIA  
"

BASE COUNT 875 a 623 c 632 g 696 t

ORIGIN

Query Match 94.1%; Score 2806.8; DB 42; Length 2826;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2811; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 21 AGCAATGGGGCCATTAAAGATTCTGTGTTCACTTGAATTTCTACCTTCTAGAAGGGC 80  
DB 1 AGCAATGGGGCCATTAAAGATTCTGTGTTCACTTGAATTTCTACCTTCTAGAAGGGC 60  
QY 81 CCTGATTAATTCATCTATTCAGCTAACAACAATGGCTATGAAGCAATTTCTGTTCAAT 140  
DB 61 CCTGATTAATTCATCTATTCAGCTAACAACAATGGCTATGAAGCAATTTCTGTTCAAT 120  
QY 141 CGACCCCAATGTCGAGAGATGAACAACACTCTCAACAATAAAGACATGGTGAACCA 200  
DB 121 CGACCCCAATGTCGAGAGATGAACAACACTCTCAACAATAAAGACATGGTGAACCA 180  
QY 201 GGCATCTCTGATCTGTTTGAAGCTACAGAAAGCATTTTATTTCAAAAATGTTGCCAT 260  
DB 181 GGCATCTCTGATCTGTTTGAAGCTACAGAAAGCATTTTATTTCAAAAATGTTGCCAT 240  
QY 261 TTGATTCCTGAAATGAGAGACAAAGGCTGACTATGTGAGACCAAACTTGAACCTA 320  
DB 241 TTGATTCCTGAAATGAGAGACAAAGGCTGACTATGTGAGACCAAACTTGAACCTA 300  
QY 321 CAAAATCTGATCTGTTGTTGCTGAGTCTACTCTCCAGTATGATGAACCTTACAC 380  
DB 301 CAAAATCTGATCTGTTGTTGCTGAGTCTACTCTCCAGTATGATGAACCTTACAC 360  
QY 381 TGAGCAGATGGGCAACTGTGAGAGAGAGGGTGAAGATCCACCTCTGATTTTCAT 440  
DB 361 TGAGCAGATGGGCAACTGTGAGAGAGAGGGTGAAGATCCACCTCTGATTTTCAT 420  
QY 441 TGAGGAAAAAGTTAGTGAATATGACCAACAAGTAGGGGATTTGCCATGATGGGC 500  
DB 421 TGAGGAAAAAGTTAGTGAATATGACCAACAAGTAGGGGATTTGCCATGATGGGC 480  
QY 501 TCATCTACGATGGGAGATTTGACAGATGACAAATATGATGAGAAATCTACTATCCAA 560  
DB 481 TCATCTACGATGGGAGATTTGACAGATGACAAATATGATGAGAAATCTACTATCCAA 540  
QY 561 TGAAGAATACAGCAGTAAAGATGTCAGAGATTAAGTGGTACAAATGATTAAGAA 620  
DB 541 TGAAGAATACAGCAGTAAAGATGTCAGAGATTAAGTGGTACAAATGATTAAGAA 600  
QY 621 GTGTGAGGAGGACGCTGTACACCAAAAGATGACATTAAGTTACAGAGCTCTA 680  
DB 601 GTGTGAGGAGGACGCTGTACACCAAAAGATGACATTAAGTTACAGAGCTCTA 660  
QY 681 TGA AAAAAGATGTGATTTGTTCTCCATCCGACAGAGGAGGCTGTATATGTT 740  
DB 661 TGA AAAAAGATGTGATTTGTTCTCCATCCGACAGAGGAGGCTGTATATGTT 720  
QY 741 TGACACACATGTTGATTTATAGTTGAATCTGTACAGACAAAACCAACAAGAAC 800  
DB 721 TGACACACATGTTGATTTATAGTTGAATCTGTACAGACAAAACCAACAAGAAC 780  
QY 801 TCCAAACACGAAAAATCAAAAATGCAATCTCCGAGACATGGGAAAGTATCCGATTC 860  
DB 781 TCCAAACACGAAAAATCAAAAATGCAATCTCCGAGACATGGGAAAGTATCCGATTC 840  
QY 861 TCGAGACTTTAAGAAACCACTCTATGACAAACAGCCACCAAAATCCCACTTCTCAT 920

DB 841 TGAAGACTTTAAGAAACCACTCTATGACAAACACCCACCAATATCCACTTCTCAT 900  
QY 921 GCTCAGATTGGACAAAGAAATGTGTGTTTACCTCTGACAAATCTGGAAGCATGGCAG 980  
DB 901 GCTCAGATTGGACAAAGAAATGTGTGTTTACCTCTGACAAATCTGGAAGCATGGCAG 960  
QY 981 TGTATACCGCTCAATGACTGAATCAAGCAGCCAGCTTTCTGCTGCAGACAGTTGA 1040  
DB 961 TGTATACCGCTCAATGACTGAATCAAGCAGCCAGCTTTCTGCTGCAGACAGTTGA 1020  
QY 1041 GCTGGGGTCTGGGTTGGGATGGTACATTTGACAGTGGCCCATGATACAAAGTAACT 1100  
DB 1021 GCTGGGGTCTGGGTTGGGATGGTACATTTGACAGTGGCCCATGATACAAAGTAACT 1080  
QY 1101 CATACAGATTAACAGTGGCAGTGAAGGAGACACACTGCGCCAAAAGATTACCTGCAGCAGC 1160  
DB 1081 CATACAGATTAACAGTGGCAGTGAAGGAGACACACTGCGCCAAAAGATTACCTGCAGCAGC 1140  
QY 1161 TTCAGAGGAGACGTCATCTGACGCGGCTTCATCGGCATTTACTGTATAGGAAGAA 1220  
DB 1141 TTCAGAGGAGACGTCATCTGACGCGGCTTCATCGGCATTTACTGTATAGGAAGAA 1200  
QY 1221 ATATCCAACTGATGATCTGAAATTTGCTGCTGACGAGATGGGGAAGACAACTATAG 1280  
DB 1201 ATATCCAACTGATGATCTGAAATTTGCTGCTGACGAGATGGGGAAGACAACTATAG 1260  
QY 1281 TGGGTGCTTTAAGAGGTCAAAAGTGTGCCATCATCCACAGTGCCTTTGGGGCC 1340  
DB 1261 TGGGTGCTTTAAGAGGTCAAAAGTGTGCCATCATCCACAGTGCCTTTGGGGCC 1320  
QY 1341 CTCTGCAAGCTCAAAAGTATAGAGGAGCTGCCAAATGACAGAGGTTTACAGCATATGC 1400  
DB 1321 CTCTGCAAGCTCAAAAGTATAGAGGAGCTGCCAAATGACAGAGGTTTACAGCATATGC 1380  
QY 1401 TTCAGATCAAGTTGACAAACATGGCTCATTTGATGATTTGGGGCCCTTTCATCAGAAA 1460  
DB 1381 TTCAGATCAAGTTGACAAACATGGCTCATTTGATGATTTGGGGCCCTTTCATCAGAAA 1440  
QY 1461 TGAAGCTGTCTCTGAGCGCTTCATCCAGCTTGAAGATTAAGGATTAACCTCCAGAACAG 1520  
DB 1441 TGAAGCTGTCTCTGAGCGCTTCATCCAGCTTGAAGATTAAGGATTAACCTCCAGAACAG 1500  
QY 1521 CCAAGTGAATGAGGACAGGATCGTGGACAGCAGCTGGGAAAGGACATTTGTTCT 1580  
DB 1501 CCAAGTGAATGAGGACAGGATCGTGGACAGCAGCTGGGAAAGGACATTTGTTCT 1560  
QY 1581 TATACCTGACAAACGACGCTCCCAAAATCTTCTCTGGGATCCAGTGGACAGAAACA 1640  
DB 1561 TATACCTGACAAACGACGCTCCCAAAATCTTCTCTGGGATCCAGTGGACAGAAACA 1620  
QY 1641 AGTGGCTTTGTAGTGAACAAAAACACAAATGECCTTACCTCCAAATCCAGGCAATGC 1700  
DB 1621 AGTGGCTTTGTAGTGAACAAAAACACAAATGECCTTACCTCCAAATCCAGGCAATGC 1680  
QY 1701 TAAAGTTGGGACTTGAATAACAGTGTGCAAGCAAGTGTCAAAACCTTGAACCTGATGT 1760  
DB 1681 TAAAGTTGGGACTTGAATAACAGTGTGCAAGCAAGTGTCAAAACCTTGAACCTGATGT 1740  
QY 1761 CACGTCCCGTGGCTCAATGCTACCTGCTCCAAATTAAGTACAGTTCGAAAGCAACA 1820  
DB 1741 CACGTCCCGTGGCTCAATGCTACCTGCTCCAAATTAAGTACAGTTCGAAAGCAACA 1800  
QY 1821 GGACACACGAAATTCACAGCCCTCTGTAGTTATGCAATATTTGCCAAGAGCCTC 1880  
DB 1801 GGACACACGAAATTCACAGCCCTCTGTAGTTATGCAATATTTGCCAAGAGCCTC 1860  
QY 1881 CCCAATCTCAGGCGCAGTGTCAACAGCCCTGATTTGAATCGTAATGGAAGAAACAGTTAC 1940  
DB 1861 CCCAATCTCAGGCGCAGTGTCAACAGCCCTGATTTGAATCGTAATGGAAGAAACAGTTAC 1920  
QY 1941 CTTGGAATCTAGTATATGAGCAGGTGCTGATGCTACTAAGATGACAGGTGCTACTC 2000

```

Db 1921 CTTGAAGTACTGATATATGAGACAGGTGCTGATGCTACTAAGATGACGGTGTACTC 1980
QY 2001 AAGGATTTTCACAACTTATGACACGAGTATGATACAGTGTAAAGTGGGGCTTGGG 2060
Db 1981 AAGGATTTTCACAACTTATGACACGAGTATGATACAGTGTAAAGTGGGGCTTGGG 2040
QY 2061 AAGGATTTTCACAACTTATGACACGAGTATGATACAGTGTAAAGTGGGGCTTGGG 2120
Db 2041 AAGGATTTTCACAACTTATGACACGAGTATGATACAGTGTAAAGTGGGGCTTGGG 2100
QY 2121 TGGCTGATTTGAGAAATGATGAAATATACATGAAATCCAGCAAGACTGAAATTAAGA 2180
Db 2101 TGGCTGATTTGAGAAATGATGAAATATACATGAAATCCAGCAAGACTGAAATTAAGA 2160
QY 2181 TGAATGTTCAACACAGCAAGTGTGTTTCAACAGACATCTCGGAGGCTCATTTGGC 2240
Db 2161 TGAATGTTCAACACAGCAAGTGTGTTTCAACAGACATCTCGGAGGCTCATTTGGC 2220
QY 2241 TTCTGATGTCACAAATGCTCCCATACCTGATCTCTCCACCGGCAATACCGACCT 2300
Db 2221 TTCTGATGTCACAAATGCTCCCATACCTGATCTCTCTCCACCGGCAATACCGACCT 2280
QY 2301 GAAGGGGAAATTCACAGGGGGGAGTCTCATTTATCTGACTTGGAGAGCTCTGGGATGA 2360
Db 2281 GAAGGGGAAATTCACAGGGGGGAGTCTCATTTATCTGACTTGGAGAGCTCTGGGATGA 2340
QY 2361 TTATGACCATGGAACAGCTCACAGATATATCATTCGAATAGTACAGATTTCTGATCT 2420
Db 2341 TTATGACCATGGAACAGCTCACAGATATATCATTCGAATAGTACAGATTTCTGATCT 2400
QY 2421 CAGAGCAAGTCAATGATCTCTTCAAGTGAATATCTACTGCTCCATCCCAAGCAAGC 2480
Db 2401 CAGAGCAAGTCAATGATCTCTTCAAGTGAATATCTACTGCTCCATCCCAAGCAAGC 2460
QY 2481 CAACTCTGAGGAAGTCTTTTGTGTTAAACAGAAACATTACTTTGAAATGGCAGAGA 2540
Db 2461 CAACTCTGAGGAAGTCTTTTGTGTTAAACAGAAACATTACTTTGAAATGGCAGAGA 2520
QY 2541 TCTTTTCATGCTATTCAGGCTGTGTATAGAGTGCATCTGAAATCAGAAATATCAACAT 2600
Db 2521 TCTTTTCATGCTATTCAGGCTGTGTATAGAGTGCATCTGAAATCAGAAATATCAACAT 2580
QY 2601 TGCAGAGATATCTTTGTTATTCCTCCACAGACTCGGCGAGAGACCTAGTCCGATGA 2660
Db 2581 TGCAGAGATATCTTTGTTATTCCTCCACAGACTCGGCGAGAGACCTAGTCCGATGA 2640
QY 2661 AAGCTCTGCTCTGCTGCTATATTCATATCAACAGACACCATCTGGCATTCACATTTT 2720
Db 2641 AAGCTCTGCTCTGCTGCTATATTCATATCAACAGACACCATCTGGCATTCACATTTT 2700
QY 2721 AAAAATATGAGAGAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2780
Db 2701 AAAAATATGAGAGAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
QY 2781 TGTGATGATTAATAATAATATATCATTCATCTTTTGTATTAATAATAATTTCTAATA 2838
Db 2761 TGTGATGATTAATAATAATATATCATTCATCTTTTGTATTAATAATAATAAAAAA 2818

```

```

RESULT 3
AB017156 2937 bp mRNA ROD 04-MAR-1999
DEFINITION Mus musculus gob-5 mRNA, complete cds.
ACCESSION AB017156
NID 93721911
VERSION AB017156.1 GI:3721911
KEYWORDS GOB-5
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.

```

TITLE Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice

JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)

MEDLINE 99160866

REFERENCE 2 (bases 1 to 2937)

AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru Komiyama, ESATO, JST, Hirohashi Cell Configuration Project: 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan (E-mail:tkom@ccp.jst.go.jp, Tel:81-298-47-7563, Fax:81-298-47-5226)

## FEATURES

## source

1..2937

/organism="Mus musculus"

/db\_xref="taxon:10090"

/cell\_type="goblet cell"

/dev\_stage="adult"

/tissue\_type="intestine"

15..2756

/gene="gob-5"

15..2756

/gene="gob-5"

/codon\_start=1

/protein\_id="BA33743.1"

/db\_xref="PID:01034712"

/db\_xref="PID:93721912"

/db\_xref="GI:3721912"

/translation="MESLSPVFLILHLLEGLVSESLIOLNNNGEIVAIADHDVPEALIOHDKMTATQASPLYLEFATGKRRFYENVALIIPESKAKREYRPLTEPKNA DVLVSTRPPLNDPERYHTICACBCKRGIHLITDPLAGKLTLOYGPDRTFYEMAH FRMGVFNENDEKFTYLSKGPQAVRCSAALITGKNVRCGGSCITNGKCVIDRTVG LYKNCFVPPPHONEKASIMFNONINSVFECEKKNQAPDQNRCLNRSTWEY IOESDEFKOTIPMTAPAPYPSLQIQRIVCVLDKSGSMLNDRNRNQSRLF LQIVYEOSVWGTAFDAAVYOSLKNLNGADRLIKHLPTYSAGSTICSGLR AFVYIKRKPPTDGESEIVLITDGEENTSSCFDVKOSGAILHTVALGPAALKEIORS KMTGLOYSDDQONNGIYNAPALSSGNALIQHSIQLESRCVNLONNMWNSVY VDSYGRDITLITWTHPITFIIMDSGVONKGLDITTKVATLOYPGTAKGFWR YSIQASQTLILVTSRASAATLPPIVTPVNNNTGKFPSPVTVASIRGASPIFK ASVVALIESVNGKATVLELDNGAGADATKNDGVYSRFTAFDANGRSVYIMLGVY TSDORAPRNPNRMYIDGWIEDGEVRNPNRPSTVYODKOLFESRSTSGSFVATN VPAAAPIDLPPOITDILKASIOGNLVNLTWAPGDYDGHGRASNYIIMRSTSYD LRDHPNHSIYVNTGILPEKASSEIEFEELGWTGPGGDTDIFAIQAVDKSNKSEY SNIRAVSYFTIAPQPIPEDSTPPCPDIDISNTIPGIVHLKIMKMLGEMOVTIGLH"

## gene

## CDS

BASE COUNT 860 a 718 c 693 g 666 t

ORIGIN

Query Match 59.1%; Score 1764; DB 12; Length 2937;

Best Local Similarity 77.5%; Pred. No. 0;

Matches 2212; Conservative 0; Mismatches 615; Indels 29; Gaps 5;

```

QY 25 ATGGGCGCATTTAAGAGTCTGTGTTCACTGATCTTCACTCTTGAAGGGCCCTG 84
Db 25 ATGGGCGCATTTAAGAGTCTGTGTTCACTGATCTTCACTCTTGAAGGGCCCTG 84
QY 15 ATGGAATCTTTAAAGAGTCTGTCTCTGATGATCCCTCTGGAAGAGATTCTG 74
Db 15 ATGGAATCTTTAAAGAGTCTGTCTCTGATGATCCCTCTGGAAGAGATTCTG 74
QY 85 AGTAATTCATCATTCACGCTGAACAACAATGGCTATGAGAGCATTTGTCATTCAG 144
Db 85 AGTAATTCATCATTCACGCTGAACAACAATGGCTATGAGAGCATTTGTCATTCAG 144
QY 75 AGTAGTCCCTCATCACAACGAGCAACAACGAGTATGAGGCGATCGTATCCCATAGAC 134
Db 75 AGTAGTCCCTCATCACAACGAGCAACAACGAGTATGAGGCGATCGTATCCCATAGAC 134
QY 145 CCCAATGTGCCGAGAGATGAAACACTCATTCACAACAATTAAGAGATGCTGACCGAGCA 204
Db 145 CCCAATGTGCCGAGAGATGAAACACTCATTCACAACAATTAAGAGATGCTGACCGAGCA 204
QY 135 CAGGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194
Db 135 CAGGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194
QY 205 TCTGTAATCTGTTGAAGCTACAGAAAGCATTTTATTTCAAAATAGTTGCCATTTTG 264
Db 205 TCTGTAATCTGTTGAAGCTACAGAAAGCATTTTATTTCAAAATAGTTGCCATTTTG 264
QY 195 TCTCATACCTTTTGAAGCTACAGAAAGCATTTTATTTCAAAATAGTTGCCATTTTG 254
Db 195 TCTCATACCTTTTGAAGCTACAGAAAGCATTTTATTTCAAAATAGTTGCCATTTTG 254
QY 265 ATTCTGAAACATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 324
Db 265 ATTCTGAAACATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 324
QY 255 ATTCCGAGAGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 314
Db 255 ATTCCGAGAGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 314
QY 325 AATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGATATGATGAACCTCACTAGAG 384

```

Db 315 AACGCTGATGCTCTGTATCAACAAACACGCTCTTAGCAATGATGAGCTTACACCGAA 374  
Qy 385 CAGATGGGCACTGTGAGAGAGAGGTGAAAGATCCACTCACTCTGATTTCAATGCA 444  
Db 375 CATATAGGAGCATGTGAGAAAAGGGATCAGATTCCACTGACTCTGACTTCTTAGA 434  
Qy 445 GGAAGAAAGTGTGATATGTGACCAAGATAGGGCACTTTGTCCATGATGGGCTCAT 504  
Db 435 GGAAGAGGCTACTCAGATGTGGCCCAAGACAGGACCTTTGTCCATGATGGGCTCAC 494  
Qy 505 CTACGATGGGAGTATTGACGAGTACAATAATGATGAGAAATTTACTTATCAATGGA 564  
Db 495 TTCCATGGGAGTGTATTATGAAATACACACGACGAGAAATTTCTACTTATCCAAAGGA 554  
Qy 565 AGAATACAGCAGTAAAGATGTCAGCAGGTATTACTGTACAAATGTAGTAAAGAGTGT 624  
Db 555 AAACCCCAAGCAGTGAAGTGTTCAGCAGGCATTTACGGTAAATATCAAGTTGTGGTGC 614  
Qy 635 CAGGAGGCACTGTTCACACCA---AGATGCACATTCATTAAGTAACGAGCTAT 681  
Db 615 CAGGAGGCACTGTTCACACCAAGTGTATATCGACAGATTAACGGGACTGTAT 674  
Qy 682 GAAAAGAGTGTGAGTGTCTCCATCCGCGCAGAGGAGAGAGGCTTCTAATGTTT 741  
Db 675 AAGACAAATGTGTATTGTTCACCAATCCACACAAACGAGAGGCTTCCATCATGTTT 734  
Qy 742 GCACAACTGTTGATTAATGATGATTAATTTCTGACAGACAAACACCAACAAAGCT 801  
Db 735 AACCAAAATATCAATTTCTGTGTTGATTTCTGACAGAAAAATACAAATCAAGAGCC 794  
Qy 802 CCAACAAAGCAAAATCAAAATGCAATCTCCGAGCAGATGGGAAGTGTCCGTATTT 861  
Db 795 CCAATGACCAAAACAGATGATCTCCGAGCAGCTGGGAAGTCTCCAGGAATCT 854  
Qy 862 GAGGACTTAAAGAAACCCCTCTATGACACAGCCCAAAATCCCACTTCATG 921  
Db 855 GAGGACTTAAAGCAACCCCTCTATGACAGCCCAAAATCCCACTTCATG 914  
Qy 922 CTGAGATTTGACAAAGATTTGTGTTTGTGCTTGAACAAATCTGGAAGCATGGCAGT 981  
Db 915 CTGCAATTTGACAAAGATTTGTGTTTGTGTTGATTAAGTCCGGGAGCATGCTGAAC 974  
Qy 982 GGTAAACCCCTCAATGATGATTAAGCAGAGCCAGCTTTTCTCTCTGACAGCATTTAG 1041  
Db 975 GATGATCTCTTAAACGATGATGATGAGCAACCGGCTTTCTCTCTGAGCTGTGAG 1034  
Qy 1042 CTGGGCTCTGGGTTGGGATGGTGCATTTGACATTTGACATGGCTGCTCAAAAGTGAATC 1101  
Db 1035 CAGGATCTCTGGGTTGGGATGGTGCATTTGACATGGCTGCTCAATGTTACAAAGCAACTC 1094  
Qy 1102 ATACAGATTAACAGTGGCAGTGCAGAGGACACACTGCCAAAGATTAACCTGACAGACT 1161  
Db 1095 AAACAGTTAAACAGTGGTGCAGAGATCTGCTGATCAACACTTACCCACAGTACT 1154  
Qy 1162 TCAGAGAGGAGCTCCATCTGACAGCGGCTTGCATGGGCTTACTGTATTAAGAAAGAA 1221  
Db 1155 GACAGAGGAGCATCTATATGCTCTGCTGCGAGCATTTACGATTAAGAAAGAG 1214  
Qy 1222 TATCAACTGATGATGATGAAATTTGCTGCTGACGAGTGGGAGAGCAACACTTAAGT 1281  
Db 1215 TATCAACTGATGATGATGAAATTTGCTGCTGACGAGTGGGAGAGCAACACTTAAAGC 1274  
Qy 1282 GGTGCTTTAAAGAGTCAAAACAAAGTGTGCTCATCTCAACACAGCTGCTTTGGGGCC 1341  
Db 1275 AGCTGCTTTGAACGTGTGAAGAGAGGGGCGCATCATCAAGAGTGGCCCTGGGAGCG 1334  
Qy 1342 TCTGACGCTCAAGAACTGAGAGGCTGTCCAAATGACAGAGGTTTACAGACATATGCT 1401  
Db 1335 GCTGCGCTTAAAGAGCTTGAAGGCTGTCCAAATGACAGAGGCTTACAGACATATCT 1394  
Qy 1402 TCAGATCAAGTTCGAAACATGGGCTCATTTGCTTTGGGGCCCTTCAATCAGAAAT 1461

Db 1395 TCGATCAGGTTCAAGAAATGCTCTGTTGATGCTTTCCAGACACTCTCAGGAAT 1454  
Qy 1462 GAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAAGGATTAACCTCCAGAACGC 1521  
Db 1455 GCGGAGATCGCTCAGCACTCCATCCAGCTGAGAGAGGAGATTAATCTCAGAAATTAAC 1514  
Qy 1522 CAGTGAATGAAATGGCAGATGATGTTGACAGCACCGTGGGAAAGGACACTTTGTTCTT 1581  
Db 1515 CAATGATGAATGGCTCAGTATGATGAGACAGCTGGGGCAAGGACACTTTGTTCTT 1574  
Qy 1582 ATCACTGGACAAAGCAGCGCTCCCAAAATCTTCTGAGATCCAGTCCAGACAGCA 1641  
Db 1575 ATCACTGGACAAAGCAGCTCTCTCAATATTTATCTGAGATCCAGCGAGTGGACAA 1634  
Qy 1642 GGTGCTTTGATGAGCAAAAACCAAAATGGCTTACCTCAAAATCCAGGACTTGT 1701  
Db 1635 AATGTTTATTAAGTAAACAAACACTAAGTGGCTTACTTCAAGTCCAGCGAGCGCT 1694  
Qy 1702 AAGTTGGCACTTGAATTAAGTGTGACAGCAAGCTCACAACCTTGACCTGCTGT 1761  
Db 1695 AAGTTGGCTTTGAAATTAAGCATTCAGCGAGCTCAGAGCTTCACTTGTACTGTC 1754  
Qy 1762 ACGTCCGCTGGCTCAATGCTACCTGCTCAATTAAGTACGACTTCCAAACGAGCAAG 1821  
Db 1755 ACGTCCGCTGGCAAGTGTCTACAGCTGCTCTTATACAGTACCCGGTATGATTAAG 1814  
Qy 1822 GACACAGCAAAATTCGCCAGCCCTGTGATTTATGCAAAATTTGCCAAGAGCCTCC 1881  
Db 1815 AACACAGGAAATTTCCCGAGCCCTGTAAACAGTGTATGCAAACTTCCGCAAGAGCCTCG 1874  
Qy 1882 CCAATTTCCAGGGCAAGTGTACAGCCCTGATTTGAATCAGTGAATGGAAAAAGCTTACC 1941  
Db 1875 CCTATTTCAGAGGCGACAGCTACAGCCCTTGAATTAATCTGGAATGGAATAACAGTAAAC 1934  
Qy 1942 TTGGAACACTGATTAATGAGAGAGTGTGATGCTACTTAAGATGAGAGGTTGTACTCA 2001  
Db 1935 CTGGAATTTACTGATTAAGGAGAGAGTGTCCGATGCCACAAAGATGATGCTGTACTCA 1994  
Qy 2002 AGTATTTACCACTTATGACAGATGATGATGATGATGATGATGATGATGATGATGAT 2061  
Db 1995 AGTATTTACCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2054  
Qy 2062 GAGTATTAAGAGGCAAGGAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 2121  
Db 2055 GAGTATTAAGAGGCAAGGAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 2114  
Qy 2122 GCGTGAATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2181  
Db 2115 GCGTGAATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2168  
Qy 2182 GATGTTCAACAAAGCAAGTGTGTTTCAAGCAAGATCTCGGAGGCTCATTTGTGCT 2241  
Db 2169 TATGTTCAAGCAAGCAAGTGTGTTTCAAGCAAGATCTCGGAGGCTCATTTGTGCT 2228  
Qy 2242 TGTGATGCTC---CAATGCTCCCATACCTGATCTCTCCACCTGSCCAATACCGAC 2298  
Db 2229 ACCAATGCTCCCGAGAGGCTCCCATTCCTGACCTTTTCCACCTTCAATATACGAC 2288  
Qy 2299 CTGAAGGAGGAAATTAACAGGAGGAGCTCATTAATTAATGATGATGATGATGATGAT 2358  
Db 2289 CTGAAGGAGGAAATTAACAGGAGGAGCTCATTAATTAATGATGATGATGATGATGAT 2348  
Qy 2359 GATTATGACATGAGAACAGCTCACAGTATATCTCTGAAATGATGATGATGATGATGAT 2418  
Db 2409 CTCAGGAGGAGCACTTCAACACTCTCACTCACTCACTCACTCACTCACTCACTCACT 2468  
Qy 2419 CTCAGGAGGAGCACTTCAACACTCTCACTCACTCACTCACTCACTCACTCACTCACT 2478  
Db 2409 CTCAGGAGGAGCACTTCAACACTCTCACTCACTCACTCACTCACTCACTCACTCACT 2468  
Qy 2479 GCCAAGCTGAGAGAGCTTTTGTAAACAGAAATTAATTTGAAATGAGCA 2538  
Db 2469 GCCAAGCTGAGAGAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2528





Db 844 ATCATGAACCTCTACTGACTTCAAAATACATCTCCATGACAGAAATGATCCACGACT 903  
 QY AATCCCACTCTCTACTGCTGACAGATGGACAAAGATGTGTGTTAGTCTTGACAAA 963  
 Db 904 CAACCTACATTTTCTATGCTCAAGTCCAAACAGGAGATGTTGTGTGTTGACTGATATA 963  
 QY 964 TCTGGAGCATGCGCATGCTGTAACCGCTCAATGACATGCAATGCAAGGACGACCTTTT 1023  
 Db 964 TCTGGAGCATGCTGTAACCGCTCAATGACATGCAATGCAAGGACGACCTTTT 1023  
 QY 1024 CTGCGGACGACAGTGGACCTGAGGCTGCTGGGTGGAGTGTGACATTTGACAGTGTCC 1083  
 Db 1024 TTGATTCAAATTTATTTAAAGGAGATCTTGTGTGGATGTTGACTTTGACAGTGTCT 1083  
 QY 1084 CATGTACAAAGTGAACATACATACATAAACAGTGGACGTGACAGGACACACTGCCAAA 1143  
 Db 1084 GAAATCCGAAATATCTAACAATAAATTAAGTATGATTAATGTAATAATATCTGCA 1143  
 QY 1144 AGATTACTGCGACGCTTCAGGAGGACGCTGATCTGAGGCGCTTCGATCGGCAATTT 1203  
 Db 1144 AATCGCTCAAGAAAGCTAATGTGGAATCTAATTTGTAAGGCGCTCAAGACAGATTT 1203  
 QY 1204 ACTGTGATGAGAGA---AATATCCAACTGATGATGATGAAATTTGCTGCTGACGAT 1260  
 Db 1204 CAGGCAATTTATTCAGAGTCAACAGACACTTCTGGTTCTGAATATCTACTAATCCGAT 1263  
 QY 1261 GGGGAGACACACTATAGTGGTGTCTTAAGAGGCTCAAGAAAGTGGTCCATCATC 1320  
 Db 1261 GGGGAGACACACTATAGTGGTGTCTTAAGAGGCTCAAGAAAGTGGTCCATCATC 1320  
 QY 1321 CACACAGTGTCTTGGGCGCTCTGACAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACA 1380  
 Db 1321 CACACAGTGTCTTGGGCGCTCTGACAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACA 1380  
 QY 1324 CACACATTCCTCTGAGGACCTTCTGCTGCAAGAAAGTGGACACTGTCAATATGACA 1383  
 Db 1381 GGAGGTTACAGACATATCTCTCAGATCAAGTTCAGAAACAAATGGCTCATTTAGCTTTT 1440  
 QY 1384 GGAGGTTACAGACATATCTCTCAGATCAAGTTCAGAAACAAATGGCTCATTTAGCTTTT 1440  
 Db 1441 GGGGCGCTTCAATCAGAAATGAGAGCTGTCTGACGCTCCATCCAGCTTGAAGTAG 1500  
 QY 1438 AGTAGAATTTCACTAGAAAGTGGACATATCTGACAGACATCTTCAAGTGGAGAGCAA 1497  
 Db 1501 GGATTAACCTCCAGAACAGCCAGTGAATGACAGTGTGCTGACAGACCCGTG 1560  
 QY 1498 GCTTGGCAATTCAGAAAAGAAATGAGTACGTTGCTGCTGATGATGATCAAT 1557  
 Db 1561 GGAAGGACACTTGTCTTATCAGCTGACAGACGACGCTCCCAATTCCTGCTG 1620  
 QY 1558 GGAATGACACTTGTCTTATGCTGATGACATGACATTAAGGACAGAAATCTTCTCCAG 1617  
 Db 1621 GATCCAGTGGACAGAGCA-----AGTGGCTTTGTAGTGGACAAAACACAAA 1671  
 QY 1618 GATCCAAAAGAAAGAAATTAACCTCGATTTCAAGAAATAGTAAATATTTAT 1677  
 Db 1672 ATGGCTTACCTCAAAATCCAGGACCTGCTAAGTGTGGCACTTGGAAATCACTCT- 1728  
 QY 1678 TCTGCTGCTTCTGAAATACCTGATATGACAGACAGTCTGTTGACTTACAGCTTTCTA 1737  
 Db 1728 -----GTAACCAAGCTCACAACCTTGACCTGCTGACGCTCCGCTGCTCAATGCT 1782  
 QY 1738 AATAATATCTCCAGCTCTCAAAATCTAAGTGTGACAGTGTGACCTGAGCAAGAAAGCT 1797  
 Db 1783 ACCCTGCTCAATTTACAGTGTCTTCAAAAAGCAAGCAAGCAACCAAAATTTCCAGC 1842  
 QY 1798 ACTACACCCCAAGTAACTGACAGCTCAATGATGATCAAAATACAGACACTACCTAGC 1857  
 Db 1843 CCTGTGATGATTTGCAAAATTTGCGCAAGAGCTCCCAATTTCTCAGGCGCAGTGC 1902  
 QY 1858 CCAATGATTTGTTATGCAAGTGTCAAGGCTTTTGTGCTGCTGCTGGAATCAATGTA 1917  
 Db 1903 ACAGCCCTGATGAAATGAGTGAAGTGAAGAAACAGTACTCTTGAATGATGATGTA 1962  
 QY 1918 ACTGCAATTTAGAACTGAAGTGTGCAATCAAGTAACTGAGGCTCTGTGGCAAAATGCT 1977

QY 1963 GCAGTGTGATGCTCTACTAAGATGACGCTGTCTACTACAGTATTTTCAACTATATGAC 2022  
 Db 1978 GCAGTGTGATGCTCTACTAAGATGACGCTGTCTACTACAGTATTTTCAACTATATGAC 2037  
 QY 2023 ACGAATGTGATATCAGTGTAAAGTGGCGCTCTGGAGGATTAACCCACGAGCG 2082  
 Db 2038 GGAATGTGATATCAGTGTAAAGTGGCGCTCTGGAGGATTAACCCACGAGCG 2097  
 QY 2083 AGATGATACCCAGAGTGGAGCTGATCATCTGCTGTGATGTGATGATGTA 2142  
 Db 2098 AGTTTAAAGACACACAGAACAAAGCCCTGTATATACAGGCTATTTAAATGTTAA 2157  
 QY 2143 ATACAAATGGAATCCACCAAGACCTGAAATTAATAGATGATTTCAACACAGAGT 2202  
 Db 2158 ATTAATGTAAGCCACACAGACCTGAAATTAATAGATGATTTCAACAGAGT 2217  
 QY 2203 TGTTCAGCAAGAACCTCTGGAGGCTCATTTGTGCTGTGATGCTCCCAATGCTCC 2262  
 Db 2218 GACTTTAGCAAGTAACTCTGTGAGGCTCATTTATGATATGAGAGCTCTCTCTGTAAT 2277  
 QY 2263 ATACCTGATCTCTCCACCTGCGCAATATCCAGCTGAAAGCGGAAATTCAGGCGGCT 2322  
 Db 2278 CACCTCTGTGTGCTCCACCAATTAATTAATAGACCTTGAACAAATTTAAAGAG-- 2336  
 QY 2323 AGTCTATTAATCTGACTTGGACAGCTCTCTGGGATGATTAATGACATGGAACAGCTCAG 2382  
 Db 2336 -ATCACAATTCAACTTATGATGACAGCCCTGCGAATGTCTTATTAAGAAAGGCAAC 2394  
 QY 2383 AAGTATATCAATGGAATGAGTATCTGATCTCAGAGCAAGTCAATGATCT 2442  
 Db 2395 AGCTACATTTAAGAAATGAGTATGATGATGATGATGATGATGATGATGATGATGAT 2454  
 QY 2443 CTTCAGTGAATCTACTGCTCTCATCCCAAGAGACCACTCTGAGAGTCTTTTGG 2502  
 Db 2455 ACTTAGTGAATCTCTCAGCTTAAACCTAAGAGGCGCGCTGATGATGATGATGATGAT 2514  
 QY 2503 TTTAACCAGAAACATTAATCTTTGAAATGGACAGATCTTTCATGCTATTCAGGCT 2562  
 Db 2515 TTTAACCAGAACTTTTGAATGAAATGGACAGATCTTTCATGCTATTCAGGCT 2574  
 QY 2563 GTTGATAGTGTGATCTGAATCAGAAATATCCAAATGTCAGAGATGCTTTGTTAT 2622  
 Db 2575 ATCATATGAGCAATCTCAGCTCAGAGTCTTAACATCGCAAGCAATCAAGTTAT 2634  
 QY 2623 CCT 2625  
 Db 2635 CCT 2637

RESULT 5  
 AF001261 3317 bp mRNA 19-NOV-1997  
 LOCUS Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF001261  
 NID 92623762  
 VERSION AF001261.1 GI:2623762  
 KEYWORDS Bos taurus.  
 SOURCE Bos taurus.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 Bovinae; Bos.  
 REFERENCE 1 (bases 1 to 3317) Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
 Eblin, R.C., Widom, J., and Paul, B.U.  
 Goodwin, A. and characterization of Lu-ECAM-1 suggest it is an  
 endothelial chloride channel  
 TITLE Unpublished  
 JOURNAL 2 (bases 1 to 3317)  
 REFERENCE Eblin, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
 and Paul, B.U.  
 AUTHORS Goodwin, A.

TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., WMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers  
SOURCE 1.3317  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="1"  
/tissue\_type="lung"  
63.2780  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="lu-ECAM-1"  
/protein\_id="AAB86529.1"  
/db\_xref="PID:92623763"  
/db\_xref="GI:2623763"  
/translation="MVLCLNVILFLHLPLGKMSVNLINNGDIVIAINPSVE  
DEKLEIKEMVTASTYLEFATRKRYERAVSLIPMTKSEYFIPROESDOAD  
VIVANPYLKYDDPTTLOYGRGCKGKIHTPNFLTNHITGSRGVYHEAHL  
RMGJFDEYNVQPYTISKNTLEATRCSTHITGNVFKCPGGSCITSLRDSQIG  
LYEAKCTFLPKRSQAKESIMFSLHSEVTECKHNTAPNLONKRCGKSTWDY  
IMNSVDPTSPTEMPNPTPTSLKSKQVVCVLDKSGMSAEDRLQOMQAAE  
LYLQVLEKSLVGMVTFDSVAEIONHLRTLDNVQKITAQLPOVANGSTICRGL  
KAGROAIHSDPTSGSEILITDGEDNEISCFEDYRSQAIIHTILGSAKELE  
TLSMTGGRYRFPAKDDITGLTNAPSRISRGSTTQOAIQLESLAKITGRKRVNGTV  
PVDSTVGNDTFVVTWIOKEPIVLDQDGRKRYTSDEKEDKLNRSLQIPGIAET  
GFTWYSLNHNASSOMLVYTTTRASPTIPVATLAMSQTHHPSPMVAQVSG  
GFLPGLSVIAIETEDHQVLELMNGAGRPADKVDGYSRFTYVNGRSLK  
VHAORNTARLNROPONKRYLVYGVYENKILINPRPVMDIAAKIEDRSRL  
SGSFYSGAPRPNHPSVPSKTDLEAFKEDYIOLSTAGNVINDKRAKSYII  
RISKSFMDRODFDNAILVNTSNLIPKAGSKENREHFRVENGKFIYSVQAIN  
EANLISEVSHVOAIKFIPLPEDSVHDGTRISBITLAIIILGPMIFSVF"

CDS

BASE COUNT 1084 a 676 c 634 g 923 t  
ORIGIN

Query Match 30.2%; Score 900.2; DB 3; Length 3317;  
Best Local Similarity 61.4%; Pred. No. 1.7e-208;  
Matches 1635; Conservative 0; Mismatches 978; Indels 48; Gaps 10;

5 TCACGGGAGATGACAGCAATGGGCGCATTAAGAGTCTGTGTCATCTGATTTCTC 64  
43 TTACTGTACATGTCGCAAAATGGTCTCTGTCTGATTTTCTGTTCTTAACCTTGC 102  
65 ACCCTTGAAGGGCCCTGAGTAATTCATCTAGCTGACGACAACTGGCTATGAG 124  
103 ATCTCTGCTGG---AATGAAGATTCATATGTAATTGATTAACTGGTATGAG 159  
125 GCATTGTGCTGCAATCGACCCCAATGTCGCAAGAGTAAGAACTCATTCACAAATA 184  
160 GCATTGTATGCAATTAACCCAGTGTGCGAAGATGAAATCAATTCATGAAACATA 219  
185 AGGAGTGTGACCCAGGATCTGTATCTGTTGAAGCTACAGAAAGCATTTTAT 244  
220 AGGAATGTATCTAGAGTTTACTTACCTGTTTACATGCGCAACAAAGAGATTAT 279  
245 TCAGAAATGTTGCCATTTTGTATCTGGAACATGGAAGCAAAAGGCTGATGTGAGC 304  
280 TCAGAAATGTGAGCAATTTAATTCGAATGACCTGGAATCAAAATCTGATCTCATAC 339  
305 CAAACTGTGAGACTACAAAATGCTGATGTTGCTGCTGAGTCTACTCTCCAGGTA 364  
340 CAAAACAGAAATCATATACACGAGGAGATGTCATAGTGTCTAATCCCTATCTAAATATG 399  
365 ATGATGACCTTACACTAGAGATGGCACTGTGGAGAGAGGTTGAAGATCCACC 424  
400 GAGATGATCCCTATACCTTCAATGTGAGAGGTGTGAGAAAAGAAATATATACATT 459  
425 TCACCTCTGATTTCAATTCAGGAGAAAAGTTAGCTGAATATGACACCAAGAGTAGGCAT 484  
460 TTACTCCAAACTTCTGTGATTAATTTCCACATCTATGGGTCCGAGGACAGATAT 519

485 TTGTCATGAGTGGGCTCATCTACAGTGGGAGTATTTGACAGTACATATATGATGAGA 544  
520 TTGTCATGAGTGGGCTCATCTCCCTGGGAAATTTTATGATATATGTGAGACAGC 579  
545 AATCTACTTATCC---AATGGAAGATATACAGCAGTAAAGTGTACACAGGATATTACG 601  
580 CATTTATATTTTCCAGAAAAGAACATATTGAAGCAACAGATGTTCAACTCATATTACTG 639  
602 GTACAAATGTAG---TAAAGAGTGTACAGGAGGAGCTGTTTACACCAAAAGATGCACAT 658  
640 GATTAATGTGTTTTCAGAAATGCTCCGAGGAGCTGATTAACAGATCATATGACAG 699  
655 TCAATTAAGTATACAGGACTCTATGAAAAAGTGTGAGTTGTTGTTCCATCCGCCGAGA 718  
700 GTGATCTACAGAGAGGCTGTATGAGCAAAATGTACATTCCTTCCAAAAATCCCGAGA 759  
719 CGGAGAGGCTCTATATGTTTGCACAACATGTGATCTATAGTTGAATTCGTACAG 778  
760 CTGCAAGGAATCCATTTATGTTATGCGCAAGCTCCATCTGTGACTGATTTTGTACAG 819  
779 AACAAACCCACAAAGAGCTCCAAACAGCAAAATCAAAATGACATCTCCGAGCA 838  
820 AAAAAACACACATACAGAGGCTCAAAACCTACAAAACAAAATGTGCAATGGCAAAAGCA 879  
839 CATGGGAAGTATCCGCTATCTGAGGACTTTAAGAAAACCACTCTATGACA-----A 892  
880 CATGGGAAGTATCTGAGGACTCTGTTGCTTCAAAATCACTCCCTGAGAGAAATGA 939  
893 CACAGCCACCAAAATCCCACTCTCATGCTGACAGATGGACAAAGAAATGTGTGTTAG 952  
940 ATCCACCCAGCTCATCTCATTTTATTTGCTCAAGTCCAAACAGGAGGTAGTCTTTGG 999  
953 TCCTTGACAAATGTGAAAGCATGGGAGCTGTACCCGCTCAATGACTGATCAAGCAG 1012  
1000 TACTGTATTAATCTGGAAGCATGTGCGAAGAGCCGCTCTTCAATTAATGAATCAAGCAG 1059  
1013 GCCACCTTTCTGCTGCGAGACAGTTGAGCTGGGCTCTGGGTTGGATGTGATATTG 1072  
1060 CAGAACTTACTGATTAAGTATTAAGAAAGGATCTTGTAGTGGATGTGATATTG 1119  
1073 ACAGTGTGCCCATATACAAAGTGAACATCATACATATTAACAGTGGAGTGACAGGACA 1132  
1120 ACAGTGTGCTGAATCCAAATCATCTACCAAGAAATACATGATATATATTTTACCAA 1179  
1133 CACTGCGCAAAAGATTTACCTGACAGAGCTTACAGGAGAGCTCATCTGACAGGAGCTTC 1192  
1180 AGATACCGCAAAAGTGTGCTGAGTGAAGTAAATGTGGAAGTCAATTTGTAGAGGCTCA 1239  
1193 GATGGCATTTACTGTGATTTAGGAAGA---AATATCCAACTGATGATCTGAATTTGTC 1249  
1240 AAGCAGGATTCGAGCAATTTCCACAGTGAACAGAGTACTCTGTGTTGTAATTCATAC 1299  
1250 TGCTGACGAGTGGGAGAGACACATATAGTGGTGTGTTTACGAGGTTCAAAACAAAGTG 1309  
1300 TATTATCTGATGGGAAATATATGAATTAATTCATGCTTTGAGGATTTAAACGAATG 1359  
1310 GTGCAATCATCCACAGTGTGCTGGGCTCTGACAGCTCAAGAGTACAGAGAGCTGT 1369  
1360 GTGCAATCATCCACAGTGTGCTGGGAGCTGTGCTGCAAGAGTGTGAGACATTGT 1419  
1370 CCAAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTTGAACAATGGCTCA 1429  
1420 CAAATATACAGGAGATATGTTTTTTCGCAATTAAGACATA-----ACTGGCTTGA 1473  
1430 TTGATGCTTTTGGGCGCTTTCATCAGGAATGGAAGTGTGCTGACGAGCTCCATCCAGC 1489  
1474 CTATGCTTTTCAATTTTCAATTTTCAATGAGTGAAGATCATCTACGAGGCTATTCAGT 1533  
1490 TTGAGAGTAAAGGATTAACCTCCAGAACAGCCAGTGTGATGAATGGCAGATGATCGTG 1549  
1534 TGGAAAGCAAGGCTTGAATAATACAGGAAGAAAGATTAACGCGACAGAGGCTGTAG 1593  
1550 ACAGCACCGTGGGAAAGACACTTTGTTTCTTATCCTGGACAAACGACGCTCCCAAA 1609





```
Db 2292 AGTAAGCTCAGACCTGGAGGCTAGTTTATAGGATTA--ATTCACTTCAATGG 2348
Oy 2344 ACAGCTCTGGGAGATGATATGACATGGACAGACACCAATATATCATTCGAATAGT 2403
Db 2349 ACGGCCCTGGCAGAGTCTGCACATGGAAGACACATAGTACATCATTCGAATAGAC 2408
Oy 2404 ACAAGTATCTTGATCTCAGAGACAGTCAATGAATCTCTTCAAGTGAATACTAGCT 2463
Db 2409 CAGCATCTCTGATGATCCAGAGATTTTAACAATGCTACTTGTGAATGCTCTCAGT 2468
Oy 2464 CTGATCCCAAGAGAGCCACTCTGAGAGCTTTTGTTTAAACCAAAACATTAAT 2523
Db 2469 CTGATACCTAAAGAGAGCTGCTCAAAAGAACATTTAAATCAACCAAACTTTTAA 2528
Oy 2524 TTGAAATGGCAGAGATCTTTCATCTGCTATTCAGGCTGTTGATAGGTCGATGANA 2583
Db 2529 ATAGCAATGGCATCCAGCTCTACATTCGAATCCAGGACAGAAATGAAGCCAGTCTCAGC 2588
Oy 2584 TCAGAAATATCCACATTCGACGAGTATCTGTTTATTCCTCCACA 2630
Db 2589 TCTGAGTCTCCACATCGCACAGCTGTCAAGCTTACTTCTCTAGA 2635

RESULT 7
AF052746 3471 bp mRNA ROD 08-APR-1999
LOCUS Mus musculus chloride channel cacc mRNA, complete cds.
DEFINITION AF052746
ACCESSION AF052746
NID 93560546
VERSION AF052746.1 GI:3560546
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3471)
AUTHORS Romo, L., Musante, L., Cinti, R., Seri, M., Moran, O., Zegarra-Moran, O.
and Galletta, L.J.V.
TITLE Characterization of a murine gene homologous to the bovine cacc
chloride channel
JOURNAL Gene 228 (1-2), 181-188 (1999)
MEDLINE 99173882
REFERENCE 2 (bases 1 to 3471)
AUTHORS Romo, L., Musante, L., Cinti, R., Moran, O., Seri, M. and
Galletta, L.J.V.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto
Gianina Gaslini, Genova T-16148, Italy
FEATURES
source location/Qualifiers
1..3471
/organism="Mus musculus"
/db_xref="taxon:10090"
/db_xref="dbest:AA245677"
47..2752
/codon_start=1
/product="chloride channel cacc"
/protein_id="AAC35003.1"
/db_xref="PID:93560547"
/translacion="MVEGLIOLFLTLHLONTSEWHLNSNGEVVIAINPSVE
DERLIPSLKEVNTQASTYLFASOGRTVFNISILVPTKMSKEPIELPKFRESIDKAD
VIVADPHLOHGDDEPTLYOGCGRGQYIHTPEFLTDNRIITGPRGRVHEHMAH
RMGVDEYNQPEYMSKNTIEATRCSTRITGNVNHRCNRCVTRACRDSKTRL
YEPKCTFIPDKIOTAGASIMEMONINSVECTEKNHNAEAPNOKNCRSTMDVI
KTSADFOVAPRMGRTEAPPTFSILKRRRVGCIYIDKSGSMKEDRLIMNOASDI
YLIOIVKESWGLVTPDSAHIONLYIKITSSSDYOKITANLPOQASGGTISGHGL
AGFOATSSDOSISGSEIVLTIDGDNIRSCFAVSSGAIHTITAGPRAAELEI
LSDMTGRLFYANDLNSLIDAFSRISSTSGVSQOALQLESKAFDVAGAWINGTVP
LDSTVGDTEFVITWVKKPEIILDDPKGKTYTSDPDDELNTRSAQLQPTAETG
TWYSITGTSQILITMTVTTRASPTMEPLATAMSGSTAYSPSMIVYARVSGEL
PVLGANTALIEAHGOVLTLEMDNGAGTAVNDGITYTTFYDHRNSRSLKVRV
OAKNRKTRLSLRKNKSLYIRGYENKIVINPRPDVQEAIEATVEDFNRVYSGGS
```

```
FTVSGAPPDGDHARVFPSPKVTYDLEAEFIDYTHLTATAPCKYLNDNRARIYIRMSO
HPILOLQFNFNATLVNASSLIPKAEAGKEFKERPEFKFIANGIQLITAIQADNEASL
TSEVSNIAQVAKMTLSLEDSISALGDISKIEATIMTIGLVIFENSLN"
BASE COUNT 1096 a 777 c 736 g 862 t
ORIGIN
Query Match 28.9%; Score 863; DB 12; Length 3471;
Best Local Similarity 61.1%; Pred. No. 1.9e-199;
Matches 1568; Conservative 0; Mismatches 960; Indels 39; Gaps 9;
Oy 91 TCACATCTAGCTCGAACAATAGCGCTATAGGCACTTGCCTGTCATGACCCCAT 150
Db 110 TCCATGCGCATCTCCACAGCAATGATACAGGCTGGTGCATTCGATTAACCCAGT 169
Oy 151 GTCCGAGAAGTGAACACCTCATTCACAAATAAAGCATGCTGACCCAGCATCTCTG 210
Db 170 GTCCCGAGAGAGAAAGGCTCATCCCAAGCATAAAGAAATGTTACTCAGCCTTACC 229
Oy 211 TATCTGTTGAAGTACAGAAAGCGATTTTATTCAAAATGTTCGATTTGATTCCT 270
Db 220 TACCTGTTGAAGCCAGCCAGGAAGATTTTATTCAAAACATAGCATATTAGTCCCG 289
Oy 271 GAACATGGAAGACAAAGCGTACATGTGAGACCAAACTTGACCTACAAAATGCT 330
Db 290 ATGACCTGGAAGTCAAAACCTGAGTACTTAATGCCAAAACAGATTCGTATGACAAAGCA 349
Oy 331 GATGTTCTGTTGCTGAGTCTACTCTCTCCAGTATGATGAACCTTACATGAGCAGATG 390
Db 350 GACGTCATAGTGGGGAGTCCACCTGACATGAGAGACGCCCTTACACCTTCAGTAT 409
Oy 391 GGCAACTGTGAGAGAGAGGTTGAAGATCCACTCTCTGATTTTATTCAGAGAAA 450
Db 410 GGACAGTGTGGGACAGAGAGATGATACATTCCTCACTTCTCTACTGAT 469
Oy 451 AAGTAGCTGAATATGACCAAGGTGGGATTTGTCATGAGTGGGTCATCTACGA 510
Db 470 AACTTGCCTATCTACGAGCCCGAGGAGATCTTTGTCCATGATGAGGCCCCATCTCCG 529
Oy 511 TGGGAGTATTGACGATGATACATATGATGAGAAATCTACTTATC--CAATGGAAGA 567
Db 530 TGGGAGTATTGATGATATGATATGATGATGATGATGATGATGATGATGATGATGAT 589
Oy 568 ATACAGCAGTAAAGTGTTCAGCAGGTATTACTGTAACAATGTATGATGATGATGATGAT 627
Db 590 ATAGAACGAAAGAGTGTTCACAGGATCTGACAGTCTGACCAATGTGTGTGACAACTGTGAG 649
Oy 628 GGAGCAGCTGTTCACCAAAAGATGATCAATTAAGTAAAGAGAGCTATGAAAGA 687
Db 650 AGAGGCACTGTGTCACAAAGGCGTGGCGGTGATCTGAAAGACAGCGCTGATGAGACC 709
Oy 688 GGATGAGTGTGTTCTCCCAATCCCGCAGAGGAGAGGCTTCTATATGTTTGACAA 747
Db 710 AAATGTACATTTATCCAGCAAAATACAGAGCGGTGGGCTCCATATGTTTCATGAA 769
Oy 748 CATGTTGATTTCTATAGTGAATCTGTACAGACAAACACACAAAGAGAGTCCAAAC 807
Db 770 AACCTTAATTTGTGGTGAATTTTGCAAAAAAACCACAAATGACAGAGCCCAAC 829
Oy 808 AAGCAAAATCAAAATGCAATCTCGAAGACATGGAAGATGCGGATCTGAGAGAC 867
Db 830 CTACAAAACAAATGTGCAATTCGCAAGAGCAGTGGGATGTATATACAGAGCTTGTCTAC 889
Oy 868 TTAAAGAAACACTCTCATG-----ACAACAGAGCCACCAAAATCCACCTTCTCATG 921
Db 890 TTTCAGATGCGCCCTCCCATAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 949
Oy 922 CTGCGATGTGAGCAAAAGATGTGTGTATCTCTTACCAATCTGAGAGATGGCAGCT 981
Db 950 CTCAAGTCCAAAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1009
Oy 992 GGTACCGCCTCATCTGAGTGAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1041
```

Db 1010 GAAGACCGTCTTATTCGAATGAATCAAGCAGCACTGTAATTAATCAATTTGTGAA 1069  
Qy 1042 CTGGGCTCTGGGTGGGATGGTGAATTTGACAGTGTCCCATATACAAATGAACTC 1101  
Db 1070 AAGGACTCTATGTGGATGTAGTACATTTGACAGCGCTGCCACATTCACAAATTTCTA 1129  
Qy 1102 ATACAGATTAAGAGTGGCAGTGCAGGACACACTCGCCAAAAGATTACTGAGAGCT 1161  
Db 1130 ATAAATAATACGATGATAGTACGACCAAAAGATACCGCAACCTCCCCAACAGGCT 1189  
Qy 1162 TCAGAGGAGAGCTCCATCTCAGCGGGCTTCGATCGGATTTACTGTGATTA---GGAAG 1218  
Db 1190 TCGGTGGAACTTCATTTGGCATGTGATCCAGGACAGAGATTTCAGCAATTAATCTCCAGT 1249  
Qy 1219 AATATTCACATGATGATGATGATTTGCTGTCTACGAGATGGGAAAGCAACATATA 1278  
Db 1250 GACCAAGCAGCTCCGTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1309  
Qy 1279 AGTGGTGTCTTAAGCAGAGTCAAAAGAGTGGTCCATCCACAGTCCGTTGGGG 1338  
Db 1310 CGTCTCTGTGGAGCGCTCTCTCGAGGGGTGCGATCCACATCCGCTTG666 1369  
Qy 1339 CCCTTCGACCTCAAGACTAGAGAGCTGTCCAAATGACAGAGATTACAGACATAT 1398  
Db 1370 CCTGGGCTGCCGAGAACTGGAGACTGTGTGGACATGACAGAGGGCTTGTTCTAT 1429  
Qy 1399 GCTTCAGATCAAGTTGAGAACAACTGGCTCATTTGATGCTTTTGGGCCCTTTCATAGA 1458  
Db 1430 GCCAACAAAGACCT-----AAACGCTTATCGATGCTTTCAGTAAATTCATCTACA 1483  
Qy 1459 AATGAGAGTGTCTCAGCGCTCCATCCAGCTGAGAGTGAAGGATTAACCCCGCAAG 1518  
Db 1484 AGTGGAGGCTCTCCAGAGAGCTCTGAGTTGGAGACAAACCTTCGATGTCAGAGCA 1543  
Qy 1519 AGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1578  
Db 1544 GGGGAGTATGATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1603  
Qy 1579 CTATTCACCTGGACAAAGCAAGCTCCCAATCTTCTCTGGATCCCATGACAGAAAG 1638  
Db 1604 GTTATCACCCTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1663  
Qy 1639 CA-----AGTGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1689  
Db 1664 TATACACCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1723  
Qy 1690 CCAGGATGCTTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1746  
Db 1724 CCGGAGCAGTGCAGAGACAGTACTTGGACTTACAGCATCAGGGGTACCAAGTCTCAGTTG 1783  
Qy 1747 TTGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1806  
Db 1784 ATTAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1843  
Qy 1807 TCCAAACGACAAAGACAGACAGCAAAATCCAGGCTCTGTGATTTATGAAATATT 1866  
Db 1844 GGTCAATAGTGCAGAGCAGAGCCAGTACCTCGGATGATGATGATGATGATGATGATGAT 1903  
Qy 1867 GCCAAGAGAGCTCCCAATTTCTCAGGAGCAGTGCAGAGCCCTGATTAATGATGATGAT 1926  
Db 1904 AGCCAAAGATTTTGGCTTTCTGGAGGACCAATGTCACAGCCCTCATAGAGTGAACAT 1963  
Qy 1927 GGAAGAAAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1986  
Db 1964 GAGACATCAAGTCACTTGGAGCTGTGGACAAATGGGGCAGTGTGATGATGATGATGAT 2023  
Qy 1987 GAGGCTGTACTCAAGTATTTCAACTTATGACACAAAGTGTGATGATGATGATGATGAT 2046  
Db 2024 GATGGATATTAACAAGATATTTACAGATTAATGAGAAATGATGATGATGATGATGATGAT 2083  
Qy 2047 GTGGGGGCTCTGGAGAGATTAAGCAGCAGAGAGATGATGATGATGATGATGATGATGAT 2106  
Db 2084 GTGGCTGTCTGAGCAGCAAGAAAGAAACCAAGCTGAGCTTAAGA---CAGAAAGAACAG 2140

Qy 2107 GCAGTGTACATACCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2166  
Db 2141 TCTTTATATATACCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2200  
Qy 2167 GAAATTAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2226  
Db 2201 GATGTCCAAAGAAAGAGCCATAGAGTACAGTGAAGAAATTTCAACAGATTAATCTGTGA 2260  
Qy 2227 GGCTCAT---TGTGGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2283  
Db 2261 GGCTGTTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2320  
Qy 2284 GGCCTAATACCGAGCTGAGAGCGGAAATTCAGCGGGGAGTCTATTAATGATGATGATGAT 2343  
Db 2321 AGTAAAGTACAGAGCTGAGAGCTGATTAAGGATTAAT---ATTACCTTACATGG 2377  
Qy 2344 ACAGTCTCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2403  
Db 2378 AGCGCCCTGGCAGAGTCTCGACAAATGAGAGACATGATGATGATGATGATGATGATGAT 2437  
Qy 2404 ACAAGTATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2463  
Db 2438 CAGCATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2497  
Qy 2464 CTATCCCAAGAAAGAGCCAACTGTAGAGAGCTTTTGTGATTAACCCAAAGAAATTAAT 2523  
Db 2498 CTGATACCTTAAGAAAGAGCTGCTCAAAAGAAATTAATTAATTAATTAATTAATTAAT 2557  
Qy 2524 TTTGAAATGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2583  
Db 2558 ATAGCAAAATGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2617  
Qy 2584 TCAGAAATATCCAACTGACAGAGTATCTTGTGATGATGATGATGATGATGATGAT 2630  
Db 2618 TCTGAGTCTCAACATGACAGAGCTGTCAAGCTTACTTCTAGA 2664

RESULT 8  
AF043976 3415 bp mRNA PRI 07-APR-1999  
LOCUS Homo sapiens CLCA homolog (hCLCA3) mRNA, complete cds.  
DEFINITION AF043976  
ACCESSION 94572288  
NID AF043976.1 GI:4572288  
KEYWORDS  
SOURCE  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 3415)  
Gruber,A.D. and Pauli,B.U.  
TITLE Molecular cloning and biochemical characterization of a truncated,  
secreted member of the human family of Ca2+-activated Cl- channels  
JOURNAL Biochim. Biophys. Acta 1444 (3), 418-423 (1999)  
MEDLINE 99196715  
REFERENCE 2 (bases 1 to 3415)  
Gruber,A.D., Elble,R.C. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1..3415  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..3415  
/gene="hCLCA3"  
19..807  
/gene="hCLCA3"  
/codon\_start=1  
/product="CLCA homolog"  
/protein\_id="AAD23734.1"  
/db\_xref="PID:94572289"









Db 651 TTGTCATGATGGGCCCATCTCCGCTGGGGAAATATTGATGATTAATTTGACACCAC 710  
 Qy 545 AATTCTACTTATTC--AATGAGAAATACAGCAGTAGATGTTGACAGGATATTACTG 601  
 Db 711 CATCTATATTTCAGAAAGAACACTATTGAAACAACAGATGTTCACTCATATTACTG 770  
 Qy 602 GTACAAATGTATA--AGAACTGTACAGGAGGACAGCTGTTACACCAAAATGACACT 658  
 Db 771 GTATTAAATGGTGTTCAGAAATGCCCTGGAGGACGCTGTAAACAACTCATCTACAGC 830  
 Qy 659 TCAATAAAGTACAGAGACTCTATGAAAAAGATGTGATGTTGTCTCCAACTCCGCCAGA 718  
 Db 831 GTGACTCACAGCAGGAGGCTGTATGAAGCAAAATGTACATTCTTCCAAAAAATCCGCA 890  
 Qy 719 CGGAGAGGCTTCTATTAATGTTTGACAACATGTTGATCTATAGTGAATTTCTGTACAG 778  
 Db 891 CTGCAAGGAATTCATTAATGTTTATGCCAAAGTCTCATCTGTGACTGAATTTTGTACG 950  
 Qy 779 AACAAAACCAACAAGAGCTCCAAACAGCAAAATCAAAATGCAATTCGCCAGCA 838  
 Db 951 AAAAAACACACATACAGAAAGCTCCAAACCTCAAAACAAATGTGCAATGGCAAAAGCA 1010  
 Qy 839 CATGGGAAGTATCCGCTGATTTCTGAGAGCTTTAGAAAAACCACTCCTATGACA-----A 892  
 Db 1011 CATGGGATGTATCATGAACTCTGTGACTTCAGAAATACATCTCCATGACAGAAATGA 1070  
 Qy 893 CACAGCCACCAAAATCCACCTTCTCATGTCTGACAGATGGACAAGAAATGTGTGTAG 952  
 Db 1071 ATCCACCACTCATCTCATTTTCTGCTCAAGTCCAAACAGGGGTAGTCTGTGG 1130  
 Qy 953 TCCTGGAATCTGAGACATGGGAGCTGTACCGGCTCAATGCAATGCAATGCAAGAG 1012  
 Db 1131 TACTGTAAATCTGGAAGCATGTCTGAGAAAGACCGTCTTTTAAATGAATCAAGAG 1190  
 Qy 1013 GCCACCTTTCTGCTGAGAGACAGTGGGGTCCCTGGGTGGATGATGATTTG 1072  
 Db 1191 CAGACTATTAATCTGATTAAGTATGAAAGGATCTTTAGTTGGATGTTACATTTG 1250  
 Qy 1073 ACAGTGTGCCATGTACAAAGTACATACATACAAAGTGGAGTGCAGGAGCA 1132  
 Db 1251 ACAGTGTGTGTAATTCAAAATCATCTAACAAAGATTAAGTATGATTTTACCAA 1310  
 Qy 1133 CACGTGCCAAAAGATTACTGTGAGAGCTTCGAGGAGGACGCTGCTGACGGGGCTTC 1192  
 Db 1311 AGATACCGCAAACTGCTCAAGTACTATGTGGAATCTCAATTTGTAGAGGGCTCA 1370  
 Qy 1193 GATCGCATTTACTGTGTATGGAAGAAATATC---CAACTGATGATCTGAAATGTGCT 1249  
 Db 1371 AAGCAGGATTCAGGCAATTAATCCAGAGTACAGTACTCTGTTCTGAAATCATAC 1430  
 Qy 1250 TGTGACGAGATGGGAGACAACTATTAAGTGGTCTTTACAGAGTCAAAACAAGTG 1309  
 Db 1431 TATTAACTGATGGGGAATATGAAATTAATTCATGCTTTGAGATGTAAACAACATGT 1490  
 Qy 1310 GTGCCATATCCACACAGTGGCTTGGGGCCCTGACGCTCAAGACTAGAGAGAGT 1369  
 Db 1491 GTGCATATATCCACCACTGTGCTGAGACCTCTGTGCTGCCAAAGACTGGAGACATTGT 1550  
 Qy 1370 CCAAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTCCAGAAATGGCTCA 1429  
 Db 1551 CAATATGACAGAGGAGATATC-----GTTTTTTGCCAAATTAAGAACTACTGGCCTTA 1604  
 Qy 1430 TTGATGCTTTTGGGGCCTTTTCATGAGAAATGAGCTGTCTCTCAGCGCTCCATCCAGC 1489  
 Db 1605 CTAAATGCTTTCAAGATTAATTCATCTAAGTGGAGAGCATCTACAGAGGCTATTTCACT 1664  
 Qy 1490 TTGAGATGAGGATTAATCCCTCCAGAACAGCGAGTGTATGAATGGAGAGATGCTGG 1549  
 Db 1665 TGAAGAGAAAGCCTTGAATTTACAGAGAGAAAGAGTAAAGGCAAGCTGCTGAG 1724  
 Qy 1550 ACAGACCGTGGGAGAGACATTTGTTCTTATCACTGTGACAAAGCAGCTCCCAAA 1609  
 Db 1725 AGAGTACAGTGGAAATACACTTTCTTTGTTGCATGTGACATACAAAAACCAAGAA 1784

Qy 1610 TCCTCTCTGGGATCCAGTGGACAGAGCA-----AGTGGCTTTGTAGTGACA 1660  
 Db 1785 TTGTTCTCCAGATCCAAAGAAAGAAATATAAACCCTGGATTTCAAGACATAGT 1844  
 Qy 1661 AAAACCCAAAATGGCCCTACTCTCCAAATCCAGGATTTGCTTAAGTTGGCCTTGGAAAT 1720  
 Db 1845 TAAATATTGATCTGCTGCTGTGCAAAATACCTGGTATTGAGAACAGATCTGGACTT 1904  
 Qy 1721 ACAGTCT-----GCAGCAAGCTCACAAACCTTGACCTGACTGTCAGTCCGCTG 1771  
 Db 1905 ACAGCTCTTAATATATCATCTCCAGCTCTCAATAGCTTAACAGTACAGTACGACTGAG 1964  
 Qy 1772 GGTCCAAATGCTACCCCTGCTCCAAATACAGTGAATTCCTCCAAAACAGAGACACAGCA 1831  
 Db 1965 CAAGAAAGTCTACTATACCCCCAGATAATGTCACAGAGCTACATGATGATCAATACAGAC 2024  
 Qy 1832 AATTCGCCAGCCCTCTGTGTATTTATGCAATATTCGCCAAGAGAGCTCCCAATTTCTCA 1891  
 Db 2025 ATTAATCTAGCCCAATGATGTTTATGCAAAAGTCAAGTCAAGGGGTTTTTGGCTTACTGG 2084  
 Qy 1892 GGGCCAGTGTACAGCCCTGATGATATCAGTGAATGGAAAAACAGTTACCTTGGAGACTAC 1951  
 Db 2085 GAATCAGTGTAAATATGCACTTTATGAACCCAGAGATGGACATCAATATGAGACTCT 2144  
 Qy 1952 TGTATATGAGACAGTGTCTGATGCTACTAAGATGACGAGTGTCTACTCAAGTATTTCA 2011  
 Db 2145 GGGCAATGTTGTCAGGCTGCTGATCTGTCAAGATGATGGCATCTACTCAAGATATCTTTA 2204  
 Qy 2012 CACTTATGACAGAAATGTTGATGATACGTGAAGAGCGGGGCTCTGGGAGATTACG 2071  
 Db 2205 CAGATTACTATGAAATGATGATGATACGTTTAAAGTACATGACAGGCAAGAAACACA 2264  
 Qy 2072 CAGCCAGACGAGAGTATACCCAGCAGAGTGGAGCACTGTATACCTGCTGTGATG 2131  
 Db 2265 CGGCTAGGCTAAATTTAAGAACACACAGAAACAAGTTCTATATGTTCCAGGCTACGTTG 2324  
 Qy 2132 AGAATGATGAATTAATGAAATGCAATCCACCAAGACTGAAATTTAATAGATGATTTAAC 2191  
 Db 2325 AAAACGGTAAATTAATCTGAAACCCACCCAGACCTGAAAGTCAAAAGTACCTGGCAAAAG 2384  
 Qy 2192 ACAAGCAAGTGTGTTGAGGAGAACATCTCGGGAGGCTATTTGTGCTGTGATGCTC 2251  
 Db 2385 CTAAATTAAGAACTTTAGCAAGTAACTCTGTGAGGCTCAATTTACTGTATCAGAGACTC 2444  
 Qy 2252 CAATGCTCCCA--TACTGTATCTTCTCCACCTGGCAATCCAGACCTGAGAGCGG 2308  
 Db 2445 CTCCTCGTGAATCACCCCTTGTGTGTTCCACCCAGTAAATTAAGATCTTGAGGCTA 2504  
 Qy 2309 AAATTCACGGGGGAGTCTATTTATCTGACTTGGACAGCTCTGGGAGATGATATGACC 2368  
 Db 2505 AGTTCAAAGAAAG--ATTATATTCAACTTTATGAGACAGCCCTCGGCAATGCTCTAGATA 2561  
 Qy 2369 ATGGACAGC 2378  
 Db 2562 AAGGAAAGC 2571

RESULT 10  
 AF001262  
 LOCUS AF001262 3288 bp mRNA 19-NOV-1997  
 DEFINITION Bos taurus clone 2 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
 complete cds.  
 ACCESSION AF001262  
 NID 92623764  
 VERSION AF001262.1 GI:2623764  
 KEYWORDS  
 SOURCE Bos taurus.  
 ORGANISM Bos taurus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 Reference 1 (bases 1 to 3288)

AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3288)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA  
FEATURES  
source Location/Qualifiers  
1..3288  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="2"  
/tissue\_type="lung"  
59..2443  
/note="endothelial adhesion molecule: chloride channel"  
/product="Lu-ECAM-1"  
/codon\_start=1  
/protein\_id="AAB86530.1"  
/db\_xref="pid:g2623765"  
/db\_xref="gi:2623765"  
translation="MVLCLNVLFLTLHLPGMKSSMVLINNGDGVIAINPSVE  
DEKLINIKEMVTEASTYLHATKRVRFRNVSILPMWKSSEFIKROSYDAD  
VIVANPYLKGDDPYLQYGRGEGKGYIHEPNFLINENFVYISRGVAFENHML  
RNGIDENVDDPEYISRKNTIEATSCNTHITGINVEKGGSGSITSLGRDSOTG  
LYEACTFLPKRSOTAKSEIMFSPSLHSTHETCTHNTPEA PNLONKCKRSMY  
LMSYDFONTSPTEMPNPRHTPSILSKKQRYVCLVLDKSSMSKAEDELPMONAE  
LLIIVIEKGSLVGVMTVDVAEIQNHLRTIDNVTQITAKLPGVANGVSTICKEL  
KAGFPAIIHSDSTSGSEILITLDEDMNSCFEDVKRSGAI IHTIALGPSANAELE  
LTSNMTGGREFRPNKIDITGLTNAFSRISRSRSGITQOALIESKALKITGRKRVGT  
PVDSTVNDTFEVTWITQKPEIVLQDPGKRYKTSDFEDKLANRSALQIPGAET  
GHWYSLINNHSSOMLTVTPTVTRASPPIPVIAAHNSOHTAHPSPMIYAYSO  
GTPYIGISVIAIITIEDGHOVTLBELMDGAGRDVYKNGITSRTFDYDNGRSLSK  
VHAOKNNRNLNKPONKVLVPEYVENGKILINPREPEVKDLAKKIDFESFLT  
SGSFVTVSGAPPNPVSPFSPSKITDLEAKEREDYIQLSWTAPGVLDKRES"

BASE COUNT 1043 a 707 c 645 g 893 t  
ORIGIN

Query Match 26.8%; Score 800.6; DB 3; Length 3288;  
Best Local Similarity 61.9%; Pred. No. 2.7e-184;  
Matches 1442; Conservative 0; Mismatches 844; Indels 45; Gaps 9;

QY 83 TGAATATTCACATCTTCACTGCTAACAACATGGCTATGAGGCAATGTCGTGCATCG 142  
DB 114 TGAAGATTCAATGTAATTTGATTAAACATGGGTATGATGGCATTGTCATTGCAATTA 173  
QY 143 ACCCCATGTGCCAGAGAATGAAACATCTCATCAACAATAAAGACATGGTACCAGG 202  
DB 174 ACCCCAGTGTGCCAGAGAATGAAACATCTCATTAACAAATGAAGAAAGTGAACGAG 233  
QY 203 CATCTCTGATCTGTTTGAAGTACAGAAAGCGATTATTTCAAAATGTTGCCATT 262  
DB 234 CTCTCTACTTACCTGTTTCAATGCCCAACGAAGAGTTTATTCAAGAAATGAGCAATTA 293  
QY 263 TGAATTCCTGAACATGGAAGACAAAGCGTACATATGTGAGACCAAAACTTGAGACTACA 322  
DB 294 TAATTCGAATGACCTGGAATCAAAATCTGAGTACTCAATACCAAAAGAAATCAATAG 353  
QY 323 AAAATCTGATGTTCTGTGAGTCTACTCTCCAGGTATGAGAACTTACACTG 382  
DB 354 ACCAGCAATGATGATGTTGCTTAATCCCATTAATAATGAGATGATGATCCCTATACAC 413  
QY 383 AGCAGATGGCAACTGTGAGAGAGAGGTAAGATCCACTCTACTCTGATTTCAATG 442  
DB 414 TTCAATATGAGAGGTGTGAGAAAGAAATATATATATTTACTCCAAATCTTGT 473  
QY 443 CAGGAAAGAGTATGATGATGACCAACAGTAGGCAATTTGTCATGATGGGCTC 502

DB 474 TGAATATATTTCCACATCTATGAGGTCCCGAGCGAGATTTGTCATGATGGGCC 533  
QY 503 ATCTAGATGGGAGATATTTGAGAGTACATATGATGAGAAATTTCTATTATCC--A 559  
DB 534 ATCTCGCGGGGAAATATTTGATGAGTATATATGAGACCGACCATTTCTATTTCCAGAA 593  
QY 560 ATGGAAGATACAGCAATGATGTTTCAGCAGGTATTTACTGGTACAAATGTAG--TAA 616  
DB 594 AGAACATATTTGAGCAAGCAAGATGTTCAATCATATTTACTGGTATTAATGTGGTTTCA 653  
QY 617 AGAAGTGAAGGAGGACGCTGTACACCAAAAGTACATTCATTAAGTAAAGTACGGAC 676  
DB 654 AGAATATGCTGGAGGAGCGCTGTATACAGATCTATGACAGCTGACACAGACAGGGC 713  
QY 677 TCTATGAAAAGGATGTGATGTTGTTCTCCAAATCCCGCAGACGAGAAAGGCTTCTATTA 736  
DB 714 TGTATGAGCAAAATGTATATTTCTCTTCCAAAATATCCAGACTGCAAGGATCCATTA 773  
QY 737 TGTTCGACAAATGTTGATTTCTATATGTTGATTTCTGTCAGAACAAACACACAAAG 796  
DB 774 TGTATTGCCAAGTCTCCATTTCTGTACTGAAATTTTGTACAGAAAACACAAATACAG 833  
QY 797 AAGCTCCAAAGCAAGCAAAATCAAAATGCAATCTCCGAGCAGATGGAAGTATCCG 856  
DB 834 AAGCTCCAAAGCTTACAAACAAATGTGCAATGGCAAAAGCACATGGATGTATATATGA 893  
QY 857 ATCTGAGGACTTTAAGAAAACACCTCTCTATGACA-----ACACAGCCACCAATCCCA 910  
DB 894 ACCTCTTGACTTCAAGATATCATCTCCATGACAGAAATGAAATCCACGAGCATCTTA 953  
QY 911 CCTTCATTTGCTGCGAATTTGACAAAGAAATGTGTGTTAGTCTTGTGACAAATTCGAA 970  
DB 954 CATTTCAATTTGCTCAAGTCCCAACACAGCGGTGTCTGTTGACTGTGATTAATCTGAA 1013  
QY 971 GATGCGAGCTGTATACCGCTCAACATGCACTGATGATGACAGGCGCTTTCCCTCTG 1030  
DB 1014 GCATGCTGAGAAAGCCGCTCTTCAATATGATATACAGACAGAACTATCTTATTC 1073  
QY 1031 AGACAGTGAAGCTGGGCTCTGGGTTGGATGGTGAATTTGACATTTGACGCTGCCATGAC 1090  
DB 1074 AAGTATGTAAGAAAGGATCTTTAGTTGGATGGTATCATTTGACAGTGTGTGAATCC 1133  
QY 1091 AAGTGAATCTATACATGATTAACAGTGGCAGTGACAGGACACACTCCGCAAAAGATTAC 1150  
DB 1134 AAATATCTATTAACAGAAATATGATGATATGTTTACCAAAAGATCACCGCAAACTGC 1193  
QY 1151 CTGACAGCTTCAGAGGAGGAGGAGCTCATCTGACAGCGGCTTGATGCGGATTTACTGTA 1210  
DB 1194 CTCAATGATCTAATGATGAGTGAATTTGATGAGAGGCTCAAAAGCAGATTCACAGCAA 1253  
QY 1211 TTAGAGAGA---AATATCCAGTATGATGATCTGAATTTGTGCTGACGAGATGGGGAAG 1267  
DB 1254 TTATCCAGATGACAGATATCTGTTGTTGGAATATATCATATTAATGATGGGGAAG 1313  
QY 1268 ACAACATTAATGAGTGGGCTCTTAACGAGCTCAACAAAGTGGTGCATCTCACACAG 1327  
DB 1314 AATATGAATTAATCATCTTTGAGAGATGTAACGAAGTGGTGAATCATCTCACACCA 1373  
QY 1328 TCGCTTTGGGGCCCTGTGAGCTCAAGAACTAGAGAGTGTCCAAAATGAGACAGGTT 1387  
DB 1374 TTGCTCTGGACCTCTGTGCGCAAAAGATGAGACATTTGCAAAATATGACAGAGGAT 1433  
QY 1388 TACAGCATATGCTTCAGATCAAGTTCAGAAACATGAGCTCATTTGCTTTGGGGCCC 1447  
DB 1434 ATC-----GTTTTTTTGGCAATTAAGACATAAGAGGCTTACTATAGTTCAGTAGAA 1487  
QY 1448 TTTTCATCAGAAATGAGAGTGTCTCTCAGCGCTCCATCCAGCTTGAGATTAAGGATTAA 1507  
DB 1488 TTTTCATCTGAAGTGTGAGACATCACTCAGCAGGCTATTCAGTTGGAAACAAAGCTTTGA 1547  
QY 1508 CCTCCAGAAACGCAAGTGAATGAGCAGAGTGTGAGTGTGAGACAGCAGCGGAGAAAG 1567  
DB 1548 AATTTACAGAAAGAAAGATTAACGGCACAGTGTCTGTGACAGATGTACAGTTGGAAATG 1607

1568 ACACCTTTGTTTCTATACCTGGACAGCAGCCTCCCAATCTCTCTGGATCCCA 1627  
1608 ACACCTTTGTTTCTATACCTGGACAGCAGCCTCCCAATCTCTCTGGATCCCA 1667  
1628 GTGACAGAGCA-----AGTGGCTTTGTAGTGGCAAAACACCAAAATGGCCT 1678  
1668 AAGGAAAGAAATATTAACCTGGATTTCAAGAGATATAGTTAAATTCATCTGCTC 1727  
1679 ACCTCCAAATCCAGCAGCTGCTAAGTTGGCAGCTTGAATACAGTCT-----GC 1729  
1728 GCTCAATACCTGATGATGAGAGAGAGTACTGAGCTTACAGCTTCTTAATATC 1787  
1730 AAGCAGCTCAAAACCTTGACCTGAGCTGACCTGACCTGACCTGACCTGACCTGAC 1789  
1788 ATGCGAGTCTCAAAATGCTAAGTACAGTACAGTACCTGACCTGACCTGACCTGAC 1847  
1790 CTCCAAATACAGTACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 1849  
1848 CCCAGTATTCAGCAGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 1907  
1850 TAGTTATGCAAAATATTCGCAAGAGCCTCCCAATCTCTGAGGCACTGACAGCCC 1909  
1908 TTGTTATGCAAGTACAGTACAGGCTTTTGCCTGATGAGGAAATCAATATAGCA 1967  
1910 TGATGATACAGTACAGTACAGGCTTTTGCCTGATGAGGAAATCAATATAGCA 1969  
1968 TTATGAAACCCAGATGACATCAAGTACATGAGTCTGAGGCACTGAGTCTGAGGCT 2027  
1970 CTGATGCTACTAGGATGAGGCTGCTACTCAAGTATTCAGGCTGAGGCACTGAGGCT 2029  
2028 GTGATACCTGCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2087  
2030 GTAGTACAGTGTAAAGTGGCGGCTGAGGAGTAAACGACCCAGACGAGAGTGA 2089  
2088 GTAGTACAGTGTAAAGTGGCGGCTGAGGAGTAAACGACCCAGACGAGAGTGA 2147  
2090 TACCCACAGAGTGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2149  
2148 GACACACAGAGTGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2207  
2150 GGAATCCACAGAGTGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2209  
2208 TGAACCCACAGAGTGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2267  
2210 GCAGAACATCTCTGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2266  
2268 GCAGAACATCTCTGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2327  
2267 CTGATCTCTCTGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2326  
2328 CTCTGATCTCTGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2384  
2327 TCATTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2377  
2385 ATATTCACTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2435

RESULT 11  
195746 878 bp DNA PAT 17-JUL-1998  
LOCUS 195746  
DEFINITION Sequence 8 from patent US 5733748.  
ACCESSION 195746  
NID 93940216  
VERSION 195746.1 GI:3940216  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 878)  
AUTHORS Yu, G. and Rosen, C.  
TITLE Colon specific genes and proteins  
JOURNAL Patent: US 5733748-A 8 31-MAR-1998;

FEATURES  
Source Location/Qualifiers  
BASE COUNT 257 a 179 c 188 g 241 t 13 others  
ORIGIN  
Query Match 26.58; Score 790.8; DB 5; Length 878;  
Best Local Similarity 97.9%; Pred. No. 6.6e-182;  
Matches 820; Conservative 1; Mismatches 15; Indels 2; Gaps 2;  
1992 TGTCTACTCAAGTATTCACACTTATGACAGCAAGTATGATGATGATGATGATGATG 2051  
1 TGTCTACTCAAGTATTCACACTTATGACAGCAAGTATGATGATGATGATGATGATG 60  
2052 GGGCTCTGGAGAGATTTAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2111  
61 GGGCTCTGGAGAGATTTAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 120  
2112 GTACATACCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2171  
121 GTACATACCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 180  
2172 TAATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2231  
181 TAATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240  
2232 ATTTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2291  
241 ATTTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
2292 CACGACCTGAGGAGGAGATTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2351  
301 CACGACCTGAGGAGGAGATTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360  
2352 TGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2411  
361 TGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
2412 TCTGATCTCAGAGACAGATTTCAATGATGATGATGATGATGATGATGATGATGATG 2471  
421 TCTGATCTCAGAGACAGATTTCAATGATGATGATGATGATGATGATGATGATGATG 480  
2472 AAAGGAGCCCACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2531  
481 AAAGGAGCCCACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
2532 TGGCAGATCTTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2591  
541 TGGCAGATCTTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
2592 ATCCAACTGCAAGAGATTTTGTATTTCTCCACAGACCTCCGACAGACCTAG 2651  
601 ATCCAACTGCAAGAGATTTTGTATTTCTCCACAGACCTCCGACAGACCTAG 660  
2652 TCCGATGAAAGCTCTGCTCTGT-CTAATATTCATATTAAGACACCATCTCTGCA 2710  
661 TCCGATGAAAGCTCTGCTCTGT-CTAATATTCATATTAAGACACCATCTCTGCA 720  
2711 TTCAATTTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2770  
721 TTCAATTTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780  
2771 GGTGAATTTTGTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2828  
781 GGTGAATTTTGTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 837

RESULT 12  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS AF039401  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.

ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Gruber A.D., Eble R.C., Ji H.L., Schreuer K.D., Fuller C.M. and Paul B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca<sup>2+</sup>-activated Cl<sup>-</sup> channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber A.D., Eble R. and Paul B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source 1. 35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1. 1617  
/gene="hCLCA1"  
1. 33522  
/gene="hCLCA1"  
767. 1265  
/rpt\_family="11 transposable element"  
/rpt\_type="dispersed"  
1587. 1596  
/gene="hCLCA1"  
/note="putative"  
1618. 1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,22427..24522,24522..26644,26644..28659,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"  
translation="MGPFKSYELLILHLBSGLNSLIOLNNNGEIVVADPNVE  
EDETLIQIKMVTQASLXLEPATKRRYFKRVALLIETWTKRADYRPRKIETKNA  
DLVAESTIPGNDPEYTFQMGKCGEKERHLPDFIAGKLAIEGPGKAFVHMAH  
LWGVETEDYNNDEKFLSNGRIQAVRCAGSAGITGVNRCOGSCYTRCFNKVTGL  
YEKGEFLVQKTEKASIMFAQHDLSIEFCTEONHNRKAPNKNOCNLSFTWEL  
RDSDEFKKTPTIOPNPTFSLIQGRIVCLVDRKSGMAIGRLNRNLNOGQFL  
LOTVELGSMGVAWTFPDSAAHOSSELIQNSGDRPLAKRPAASGGSTSCGRSA  
FTVIRAKRPTGSESEIVLFDGEDNITSCSNVRSQALIRHVAGSPRAQDELTSLK  
MTGLQITASDQVONNGLIDAGALSNGAASORSIDLESGLTQNSQWNGVIV  
DSYVKDRLFLITWTQPOILMDPSQKOGFVVDNKTAKAYLIQIPGAKVGMKY  
SLQASQTLTLIVTSRASNAIPLPITVTSKTKDTSKFPSPILVVAIRQGSPLIRA  
SVTALIESVNGKTVLELLDNGADATKDDVYSRYFTYDNGRYSVKVALAGVN  
AARRVITPOSGALYIPGMINDLEIOMPRREINKDDVOHROVFEFSTSGSGSPVAS  
DVNAPIPDLPPGOITDLNMAIHGSLINTLWTAPGDDYHGTGTHAKTIIRISTILD

LRDKNESLQVNTALLIPKREANSEVFLFKPENITFENGTDLFIQAVDKVLSKEI  
SNIAKVSLEIPQIPPEPSPDETSAPCPNHIINSTIPGITHIKMMRWIGLQLSIA  
" 6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 19.9% Score 594.4; DB 11; Length 35278;  
Best Local Similarity 99.7%; Pred. No. 2,9e-134;  
Matches 606; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
QY 2376 AGCTCACAAGTATATCATTCGATAGTACAAAGTATTTGATCTCAGACAGTCAA 2435  
|||||  
DB 32917 AGCTCACAAGTATATCATTCGATAGTACAAAGTATTTGATCTCAGACAGTCAA 32976  
QY 2436 TGAATCTCTTCAAGTATCTACTGCTCTCATCCCAAGGAGCCACTGAGGAGT 2495  
|||||  
DB 32977 TGAATCTCTTCAAGTATCTACTGCTCTCATCCCAAGGAGCCACTGAGGAGT 33036  
QY 2496 CTTTGTGTTAAACGAAACATTAATCTTTGAAATGGCAGATCTTTGATGCTAT 2555  
|||||  
DB 33037 CTTTGTGTTAAACGAAACATTAATCTTTGAAATGGCAGATCTTTGATGCTAT 33096  
QY 2556 TCAGGCTGTTGATAGTCTGATCTGAAATCAGAAATATCCAACTTGACGAGTACTTT 2615  
|||||  
DB 33097 TCAGGCTGTTGATAGTCTGATCTGAAATCAGAAATATCCAACTTGACGAGTACTTT 33156  
QY 2616 GTTTATCTCTCAGACACTCCGCCAGACACTGCTGTATGAAAGCTCTGCTCTTG 2675  
|||||  
DB 33157 GTTTATCTCTCAGACACTCCGCCAGACACTGCTGTATGAAAGCTCTGCTCTTG 33216  
QY 2676 TCCTATATTCATATCAACAGACCACTTCCTGGCATTCATTTAAATAATATGAGAA 2735  
|||||  
DB 33217 TCCTATATTCATATCAACAGACCACTTCCTGGCATTCATTTAAATAATATGAGAA 33276



QY	1926	AGATTGGACAAAGAAATTGTGGTTTATCTCTTACAAAACTCTGGAAAGCATGGCGACTGGTA	985
Db	958	AGGCTGGGACAAAGATGGTCTGTTTATGTGTGATGTGTTCACAGAAATGGAGAGGCTG	1017
QY	986	ACCGCTTCATTCGACTCAATCAACGACGAGCGCTTTCTCGTGCAGACATGGTAGCTGG	1045
Db	1018	ACAGACTCTCTTCACTACATCAACAAAGCCGCAAAATTTATTTGATGCAGATGTGTGAATTC	1077
QY	1046	GGTCCTGGGTTGGGATGGTGACATTTGACAGTGTCTCCCATGTACAAAGTCAATAC	1105
Db	1078	ATACCTTGTGGGATTTGCCAGTTTGCACAGCAAAAGAGATATCAGAGCCAGCTACACC	1137
QY	1106	AGATTAACAGTGGCAGTGACAGGACACACTGCCAAAAGATTAC-----CTGCAGAG	1155
Db	1138	AAATTAACAGGAATGATGATGAAAGTTGCTGTTTCATATCTGCCACCACCTGTATAG	1197
QY	1160	CTTCAGAGGAGACGCTCCATCTGCGACGCGGCTTGATCGGCATTTACTGTGATTAGAA--	1218
Db	1198	CTAAACACGATATAGCATTTGTTTCAGGCGTTTAAGAAAGATTTGAGGTGGTTGAATAAC	1257
QY	1218	-GAAATATCCAATGATGATCTGAAATTGTGCTGCTGACGGATGGGGAAGACAACACTA	1276
Db	1258	TGAATGCAAAAGCTTATGGCTCTGTGATGATTTAGTGACACAGCGAGATGATTAAGCTTC	1317
QY	1277	TAAATGGGTGCTTTAACGAGTGTCAAACAAAGTGTGCCATATCCACACAGTGTCTGG	1336
Db	1318	TTGGCAATTTGCTTTCACCACTGTCTCAGCAAGGTGTTCAACAAATTCATCTCATTTGCCCTGG	1377
QY	1337	GGCCCTCGCAGCTCAGACGACTAGAGAGGTGTCCAAATAGACAGAGAGTTTACAGACAT	1396
Db	1378	GTTATCTGCAGCCCCCAATCTGAGAGAAATTATCAGCTTTACAGAGAGTTTAAAGTTCT	1437
QY	1397	ATGCTTCAATCAAGTTTCAGAACAAATGGCCCTAATTGATGCTTTTGGGCGCTTTCATCAG	1456
Db	1438	TTGTTCAGATATATCAAACTCCAAATGATGATGATGATCTTTCAGTAGAATTTCTCTCG	1497
QY	1457	GAATGAGAGCTGTCTCAGCGCTCCATCCAGTTTAGAGTAAAGGATTAACCTTCAGA	1516
Db	1498	GAACTGGAGACATTTTCCACCAACATATTCAGTTTAAAGTACAGGTGAAATGTGCAAC	1557
QY	1517	ACAGCAGTGGATGAATGGCAGAGTAGTCTGTGACAGCAGCGGGGAAGAGCACTTTGT	1576
Db	1558	CTCACCAATCAATTAATAAAACACAGCTACTGTGATATATCTGTGGCACAACCAATATGT	1617
QY	1577	TTCTTATCACCTG--GACAAGCAGAGCTTCCCAAAATCTTCTGTGGATCCAGTGCAC	1633
Db	1618	TTTAGTACGTGGCAGCGCAGAGTGTCTCTGAGATATATATTATTTGATCTGTATGAC	1677
QY	1634	AGAAACA-----AGTGTGCTTTGTAGTGAACAAAAACACCAAAATGCGTCACTTCAAA	1687
Db	1678	GAAATATCTATCACAAATATATTTATTCACCATCTTAATCTTTTGGAGAGCTATGTTTGA	1737
QY	1688	TCCAGGATTTGTAAGTGTGGGCACTTGGAAATACAGTGTGCAAGCAAGCT-----	1739
Db	1738	TTTCAGGAACGCTAACCTCTGGGCACTGGACTTTCACCTGAAACAATATCCATCTTCTC	1797
QY	1739	CACAAACCTTTCACCTCACTGTACAGTCCCGTGCCTGCATATGCTACCTGCTCCATTA	1798
Db	1798	TGCAAGCGCTGAAGTATGACAGTAGCTGTGGCGCTTCGAACATGAGTGTGCCCCCAGCA	1857
QY	1799	CAGTGAATCTCCAAAACGACACAGGACACGACAAATTTCCCAAGCCCTGTGTAATTATG	1858
Db	1858	CTGTGGAGAGCTTTGTGTGAAGAAGACAGCCTCATTTTCTCATCTCTGTAGATTTATG	1917
QY	1859	CAATATTTGCGCAAGAGAGCTTCCCAATTTCTCAGGGCCAGTGTACACAGCCCTGATTAAT	1918
Db	1918	CCAATGTGAACAGGGAATTTATTCATCTTATATGCCACTGTCACTGCACAGTTGTAGC	1977
QY	1919	CAGTGAATGCAAAACAGTATACCTTGGATCTATGATATGAGAGAGGTGTGATGCTA	1978
Db	1978	CAGAGACTGAGAGATCCGTGTTACGCTAGACACTCTTATATATGAGAGAGTGTGATGTGA	2037
QY	1979	CTAAGATGACGCTGTCTACTCAAGGATATTTCAAACTTATGACAGAAATGATAGATACA	2038

Dd	2038	TAAAAATGATGGATTACTTCGAGGTATTTTCTCCTTGCTGCAGAAATGGTAGAATA	2097
Oy	2039	GTTGAATATGCGGGCTCTGGGAGAGTTAACGCCAGCCAGACGGAGAGTATACCAGC	2098
Dd	2098	GCTTAAGAATGCAATGTCATCATCTCCACAGATTAAGACCCCAGCCACTCATTTCCAG	2157
Oy	2099	AGAGTGGAGCACTGTACATACCTGGCTGGATTGGAATGATGAATATCATGCAATCCAC	2158
Dd	2158	GGAGTCATGCTATGATGTATACAGGTTACACAGCAAACGGTAAATATTCAGATGATCTC	2217
Oy	2159	CAAGCACGTGAATTAATTAAGATGATGTTCACACACAAGCAAGTGTTCAGCAGAACT	2218
Dd	2218	CAGGAAAATCAGTAGGCCAGAAATGGAGAGCCAAAG---TGGGCTTTTACCGAGTCA	2274
Oy	2219	CCTGGGAGGCTCATTTGTGGCTTCTGATGTGCCAAATGCTCCATACCTGATCTTCC	2278
Dd	2275	GCTCAGAGAGGCTCCTTTCACTGCTGGGAGTTTCCAGCTGGGCCCCACCTGATGTGTTCC	2334
Oy	2279	CACCTGGCCAAATCACCGCACTGAAAGCGGAATTCACGGGGCAGTCTATTAATCTGA	2338
Dd	2335	CACCATGCAAAATTAATGATTCCTGGAAAGCTGTAA--AAGTAGAAGAGAAATTGACCTAT	2391
Oy	2339	CTTGACAGCTCCGGGGATGATATGACATGCAATGGACACAGCTCAACATATATCATTTGAA	2398
Dd	2392	CTTGACAGCACTCTGGAGAAAGACTTGTATCGGGCCAGGCTACAGCTATGAATTAAGAA	2451
Oy	2399	TAAAGTCAAGATTTCTTGATCTCAGAGCAAGTTCATGATGATCTTCAAGTGAATCTA	2458
Dd	2452	TGAGTAAAGTCTACAGATATCCAAGATGACTTTTAACATGCTATTTGTAATATCAT	2511
Oy	2459	CTGCTCTCATCCCAAGAGCAACTCTGTGAGAGTCTTTTGTAAACCAAGAAAAACA	2518
Dd	2512	CAAAAGCGAAATCCACAGAGAGCTGGCATCAGGGAGATATTACCTTCTCACCCCAATTT	2571
Oy	2519	TTACTTTGGAATG 2533	
Dd	2572	CCACGAATGACCTG 2586	
RESULT 14			
AF001264			
LOCUS	AF001264	1265 bp	mRNA
DEFINITION	Bos taurus clone 4 endothelial adhesion molecule Lu-ECAM-1 mRNA,		NM
	complete cds.		19-NOV-1997
ACCESSION	AF001264		
NID	92623768		
VERSION	AF001264.1	GI:2623768	
KEYWORDS			
SOURCE			
ORGANISM	Bos taurus.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;		
	Bovinae; Bos.		
REFERENCE	1 (bases 1 to 1265)		
AUTHORS	Edble,R.C., Widom,J., Gruber,A.D., Abdel-g hany,M., Levine,R.,		
	Goodwin,A. and Pauli,B.U.		
TITLE	Cloning and characterization of Lu-ECAM-1 suggest it is an		
	endothelial chloride channel		
JOURNML	Unpublished		
REFERENCE	2 (bases 1 to 1265)		
AUTHORS	Edble,R.C., Widom,J., Gruber,A.D., Abdel-g hany,M., Levine,R.,		
	Goodwin,A. and Pauli,B.U.		
TITLE	Direct Submission		
JOURNML	Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,		
	Ithaca, NY 14853, USA		
FEATURES	Location/Qualifiers		
source	1..1265		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/tissue_type="lung"		
	/cell_type="aortic endothelial cells"		
	/clone="4"		



**CDS**

109. .1137  
/note="endothelial adhesion molecule; chloride channel;  
secreted form"

BASE COUNT	440 a	231 c	231 g	363 t
ORIGIN				

Query Match	Score	DB 3;	Length
12.88;	381;		1265;
64.38;			
64.38;			

	Matches	64.0:	Conservative	0:	Mismatches	340.0:	Indels	15:	Gaps	4:
QY	5	TCACAGGAGATGTACACGAATGGGGCCATTAAAGATTCTGTGTTCAATCTTGATTTCTTC								64
Db	89	TTACGTAAACATGTGCAAAATATGGTGTCTGTCTGTAATGTATTTCTGTCTTAACCTTGC								148
QY	65	ACCTCTAGAAAGGGGCCCTGAGTAATTCACCTCATTCAGCTGAACAACATGGCTATAGAG								124
Db	149	ATCTTTTGGCTGG---AATGAAAAGTTCATGGTAATTTGATTAACAATGGGTATGAGG								205
QY	125	GCATTGTGCTTGCATATCGACCAATATGTCCAGAAAGATGAACACATTCATCAACAATTA								184
Db	206	GCATTGTCACTTCGATTAATCCCAAGTGTCCAGAAAGATGAACAACTCATTTGAAAAACATA								265
QY	185	AGGACATGGTGACCCAGGCGATCTGTGATCTGTTTGAAAGCTACAGAAACGATTTTAT								244
Db	266	AGGAATAGGTAAGTAAAGCTTCTACTTACTCTGTTCATCTCCACAAACGAAGGTTAT								325
QY	245	TCAAAATATGTTCCCATTTGATTTCTCTGAAACATGGAAGCAAAAGGCTGACTATGTGAGC								304
Db	326	TCAGGAATGTGAGCAATTTTAATTCATACCTCGAAATCAAAATCTGTAGTACTTCATAC								385
QY	305	CAAACTTGAGACGTACACAAAATGCGATGTTCTGGTGTCTGAGTACTCTCTCCAGGTA								364
Db	386	CAAAACGAAGATCATATATGACCCAGGCGAGATGTATATGTTCTTAATCTCTTAATTAATG								445
QY	365	ATGATGAACCCCTACACTGAGATGGCGACACTGTGSAAGAAAGGTTGAAGGATCCACC								424
Db	446	GAGATGATCCCATATACACTCAATATGAAAGGTGTGSAAAAAAGSAAATATATTCATT								505
QY	425	TCACTCTCTGATTTCAITTCGAGSAAAAAGTTAGCTGAATATGAGCCACAGGTAGGGCAT								484
Db	506	TTACTCCAAACTCTTGTTGACTTAATTAATTCACACATCTATGGGTCCGAGGACAGATAT								565
QY	485	TTGTGTCATGAGTGGGGCTCATCTACGATGGGGAGTATTTTACAGATACATTAATGAGAA								544
Db	566	TTGTCCATGAGTGGGCCCATCTCCGCTGGGGATATTTGATATAGTATATGTGGACCAC								625
QY	545	AATCTACTATTATC---AATGGAAGATTCACAGCATAGATGTCAGAGGATATTACG								601
Db	626	CATTCTATATTTCCAGAAAAGAACACTATTGAAAGCACAAAGATGTTCACTATATTACG								685
QY	602	GTACAAATAGTAG--TAAAGAGTGTCAAGGAGGACGCTGTTACACCAAAAGATGACAT								658
Db	686	GTATTAATGTGTTTTCAGAAATATGCCCGAGGACGAGCTGTAACAACTATATGAGAGC								745
QY	659	TCAATTAAGTAACAGAGACTATGAAAAAGATGTGAGTTGTCTCCAAATCCGCCAGA								718
Db	746	GTGACTTCACAGACGGCGTATGAAGCAAAATGTACATTTCTCTCAAAAATATCCAGA								805
QY	719	CGGAGAGAGCTTCTTAATGTTGACACAAATGTTGATTTCTATGTTGAATTTCTGACG								778

Db	806	CTGAAAGGAAATCCATTATATGTTTATGCGAAGTCTCCATCTGTGAGACTGAATTTTGTACAG	865
Qy	779	AACAAAACCCACACAAAAGAGCTCCAAACAGCAAAATCAAATAATCAATTTCCGAGCA	838
Db	866	AAAAAACACACAAATACGAAAGCTCCAAACCTACAAAAACAAATGTGCAATGGCCAAAAGCA	925
Qy	839	CATGGGAAGTATCCGATGTTCTGAGAGCTTTAGAAAAACACTCCTATGACA-----A	892
Db	926	CATGGAGATGTAATCATATAACTCTGTGTGACTTTTAGAATTCATCTCCCATGACAGAAATGA	985
Qy	893	CACAGCCACCAAAATCCCACTCTCCATTTGCTGCAGATTGGACAAAGAATGTGTGTTAG	952
Db	986	ATCCACCGACTCACTACATTTTCATTTGCTCAAGTCCAAACAGCGGAGTAGTGTGTGG	1046
Qy	953	TCCTTGACAAATCTGGAAGCATGCGCAGCTGTAAAC	987
Db	1046	TACTTGATTAATCTGGAAGCATCTCTGCGCAACAC	1080

RESULT	15							PAT	23-DEC-1997
LOCUS	166494/c								
DEFINITION	Sequence 14 from patent US 5670367.								
ACCESSION	U66494								
MID	G2724471								
VERSION	166494.1								
KEYWORDS	GI:2724471								
SOURCE	Unknown.								
ORGANISM	Unknown.								
REFERENCE	Unclassified.								
AUTHORS	1 (bases 1 to 7218)								
TITLE	Dorner,F., Schefflinger,F. and Falkner,F.Gunter.								
JOURNAL	Recombinant fowlpox virus								
FEATURES	Patent: US 5670367-A 14 23-SEP-1997;								
SOURCE	location/qualifiers 1..7218								
BASE COUNT	/organism-"unknown"								
ORIGIN	1944 a 1491 c 1486 g 1929 t 368 others								

Query Match	Similarity	1.7%;	Score 52;	DB 5;	Length 7218;
Best Local	Similarity	4.9%;	Pred. No. 0.015;		
Matches	19;	Conservative	211;	Mismatches 156;	Indels 0;
				Gaps	0.
QY	442	GCAGAAAAAGTTAGTCGATATGACCCACAGGTAGGCGATTGTCATGAGTGGGCT	501		
DB	1463	GTAGTTAAAGATGTGGAAGATTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRR	1404		
QY	502	CATCAGATGGGGGATTTGACAGATACATATGATGAGAAATTTCTACTATCCAA	561		
DB	1403	RR	1344		
QY	562	GGAGAAATACAGCAGTAAAGATGTTCAACAGGATTTACTGTGCAAAATGTAAGAA	621		
DB	1343	RR	1284		
QY	622	TGTCAGGAGGCGAGCTGTGTACACCAAGATGCACATTCAAATAAAGTAACAGACT	681		
DB	1283	RR	1224		
QY	682	GA AAAAGATGTGAGTTGTTCTCCAAATCCCGCAGACGAGAAAGCTTCTTAATGTT	741		
DB	1223	RR	1164		
QY	742	GCACACATGTTGATTTATAGTGAATTCGTGTACAGAAACACACACAGAAGCT	801		
DB	1163	RR	1104		
QY	802	CCAAACAGCAAAATCAAAAATGCA	827		
DB	1103	RR	1078		



Mon Aug 9 13:26:14 1999

us-09-049-696-20\_1.rge

Page 23

Search completed: August 5, 1999, 23:27:21  
Job time: 8524 sec

---

**This Page Blank (uspto)**



```

OY 1052 TGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAAATAGTACAGAT 1111
DB 361 TGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAAATAGTACAGAT 420
OY 1112 TCTGATCTCAGAGCAAGTTCATGAAATCTCTCAGTGAATACCTAGCTCTATCCC 1171
DB 421 TCTTATCTCAGAGCAAGTTCATGAAATCTCTCAGTGAATACCTAGCTCTATCCC 480
OY 1172 AAAGGAGCCAACTCTGAGGAAGTCTTTTGTAAACGAGAAACATTAATCTTTGAAA 1231
DB 481 AAAGGAGCCAACTCTGAGGAAGTCTTTTGTAAACGAGAAACATTAATCTTTGAAA 540
OY 1232 TGGCAGACATCTTTTATTCCTATTCAGGCTGTGTAAGGTCGATCGAAATCGAAT 1291
DB 541 TGGCAGACATCTTTTATTCCTATTCAGGCTGTGTAAGGTCGATCGAAATCGAAT 600
OY 1292 ATCCAACTATGACAGATCTTTTGTATTCCTCAGAGCTCGGCGAGAGACCTAG 1351
DB 601 ATCCAACTATGACAGATCTTTTGTATTCCTCAGAGCTCGGCGAGAGACCTAG 660
OY 1352 TCCGTATGAAAGCTGCTCTCTTGT-CCTAATATTCATATCAACAGACACATTCCTGCA 1410
DB 661 TCCGTATGAAAGCTGCTCTCTTGT-CCTAATATTCATATCAACAGACACATTCCTGCA 720
OY 1411 TTCACATTTTAAATATATGTAAGTGGATAGGAGACTGCACTGTCAATAGCTAG 1470
DB 721 TTCACATTTTAAATATATGTAAGTGGATAGGAGACTGCACTGTCAATAGCTAG 780
OY 1471 GCTGAATTTTGTGAGATTAATTAATTAATCATTCATCTTTTGTGATTAATA 1528
DB 781 GCTGAATTTTGTGCGGTGAAT-AAATATATSAITTCANCCITTTTGTGATTAATA 837

RESULT 2
ID V16672 standard; cDNA; 878 BP.
AC V16672.
DT 22-JUN-1998 (first entry)
DE Polynucleotide sequence of a colon-specific gene.
KW Colon-specific gene; probe; detection; expression; human;
KW diagnostic assay; colon cancer; antibody; screening; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 2..685
FT /tag= a
FT /note= "no stop codon given"

US5733748-A.
PD 31-MAR-1998.
PF 06-JUN-1995; 469667.
PR 06-JUN-1995; US-469667.
PA (HOMA-) HUMAN GENOME SCI INC.
PI Rosen C, Yu G;
PI WPI: 98-229823/20.
DR P-PSDB: W46879.
PT Colon-specific nucleic acids - useful as probes for detecting colon
PT cancer micrometastases
PS Claim 15; Fig 5A-B; 51pp; English.
CC V16668-81 represent polynucleotide sequences of partial or full-length
CC cDNA clones of colon-specific genes. The polynucleotides can be used
CC as probes to detect expression of the corresponding human genes; e.g. in
CC diagnostic assays for detecting micrometastases of colon cancer.
CC Recombinant cells containing the polynucleotides can be used to
CC produce the polypeptides, in order that antibodies can be raised and
CC used in further screening or diagnostics.
SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 47.0%; Score 790.8; DB 1; Length 878;
Best Local Similarity 97.9%; Pred. No. 5.6e-212;
Matches 820; Conservative 1; Mismatches 15; Indels 2; Gaps 2;
692 TGTCTACTCAGAGTATTTACAACTTATGACACGAGATGTAATAGTGAATGAAAGTCCG 751

```

```

DB 1 TGTCTACTCAGAGTATTTACAACTTATGACACGAGATGTAATAGTGAATGAAAGTCCG 60
OY 752 GGCCTCTGGAGAGAGTTTAAACGACAGCAGAGAGTGTATACCCAGCAGAGTGGAGCACT 811
DB 61 GGCCTCTGGAGAGAGTTTAAACGACAGCAGAGAGTGTATACCCAGCAGAGTGGAGCACT 120
OY 812 GTACATACCTGGCTGGATTTGAGATGATGAATATCAATATGAAATCCACAGACCTGAAAT 871
DB 121 GTACATACCTGGCTGGATTTGAGATGATGAATATCAATATGAAATCCACAGACCTGAAAT 180
OY 872 TAATAGAGTATGTTTCAACAACGAAAGTGTGTTTTCAGAGAACATCTCGGGAGGCTC 931
DB 181 TAATAGAGTATGTTTCAACAACGAAAGTGTGTTTTCAGAGAACATCTCGGGAGGCTC 240
OY 932 ATTTGTGCTTCTGATGTGCCAAATGTCTCCATACCTGATCTCTCCACCTGGCCAAAT 991
DB 241 ATTTGTGCTTCTGATGTGCCAAATGTCTCCATACCTGATCTCTCTCCACCTGGCCAAAT 300
OY 992 CACGACCTGAAGGCGGAATTCACGGGGGCACTCTATTAATCTGACTTGGACAGCTCC 1051
DB 301 CACGACCTGAAGGCGGAATTCACGGGGGCACTCTATTAATCTGACTTGGACAGCTCC 360
OY 1052 TGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAAATAGTACAGAT 1111
DB 361 TGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAAATAGTACAGAT 120
OY 1112 TCTGATCTCAGAGCAAGTTCATGAAATCTCTCAGTGAATACCTAGCTCTATCCC 1171
DB 421 TCTTATCTCAGAGCAAGTTCATGAAATCTCTCAGTGAATACCTAGCTCTATCCC 480
OY 1172 AAAGGAGCCAACTCTGAGGAAGTCTTTTGTAAACGAGAAACATTAATCTTTGAAA 1231
DB 481 AAAGGAGCCAACTCTGAGGAAGTCTTTTGTAAACGAGAAACATTAATCTTTGAAA 540
OY 1232 TGGCAGACATCTTTTATTCCTATTCAGGCTGTGTAAGGTCGATCGAAATCGAAT 1291
DB 541 TGGCAGACATCTTTTATTCCTATTCAGGCTGTGTAAGGTCGATCGAAATCGAAT 600
OY 1292 ATCCAACTATGACAGATCTTTTGTATTCCTCAGAGCTCGGCGAGAGACCTAG 1351
DB 601 ATCCAACTATGACAGATCTTTTGTATTCCTCAGAGCTCGGCGAGAGACCTAG 660
OY 1352 TCCGTATGAAAGCTGCTCTCTTGT-CCTAATATTCATATCAACAGACACATTCCTGCA 1410
DB 661 TCCGTATGAAAGCTGCTCTCTTGT-CCTAATATTCATATCAACAGACACATTCCTGCA 720
OY 1411 TTCACATTTTAAATATATGTAAGTGGATAGGAGACTGCACTGTCAATAGCTAG 1470
DB 721 TTCACATTTTAAATATATGTAAGTGGATAGGAGACTGCACTGTCAATAGCTAG 780
OY 1471 GCTGAATTTTGTGAGATTAATTAATTAATCATTCATCTTTTGTGATTAATA 1528
DB 781 GCTGAATTTTGTGCGGTGAAT-AAATATATSAITTCANCCITTTTGTGATTAATA 837

RESULT 3
ID T22483 standard; cDNA to mRNA; 255 BP.
AC T22483;
DT 22-AUG-1996 (first entry)
DE Human gene signature HUM504094.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W0951472-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
PI WPI: 95-206931/27.

```

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in T19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SQ Sequence 255 BP: 76 A; 52 C; 39 G; 80 T;

Query Match 14.3%; Score 241.2; DB 1; Length 255;  
 Best Local Similarity 95.7%; Pred. No. 2.7e-58;  
 Matches 243; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 1275 GATCTGAATTCAGAAATATCAACATTCGACGAGTATCTTTGTTATTCCTCCACAGACT 1334  
 DB 1 GATCTGAATTCAGAAATATCAACATTCGACGAGTATCTTTGTTATTCCTCCACAGACT 60  
 OY 1335 CCGCAGAGACACCTTATGCTGATGAACGCTGCTCTGCTTCTTAATTCATCAAC 1394  
 DB 61 CCGCAGAGACACCTTATGCTGATGAACGCTGCTCTGCTTCTTAATTCATCAAC 120  
 OY 1395 ACACACATTCCTGGCATTCACATTTAAATATGTAAGAGTAGAGACGCGAG 1454  
 DB 121 ACACACATTCCTGGCATTCACATTTAAATATGTAAGAGTAGAGACGCGAG 180  
 OY 1455 CCGTCAATGAGCTAGGCTGAAATTTTGTACATTAATTAATTCATTCCTTT 1514  
 DB 181 CCGTCAATGAGCTAGGCTGAAATTTTGTACATTAATTAATTCATTCCTTT 240  
 OY 1515 TTTTGATTATATAA 1528  
 DB 241 TTTTNGTTTATAA 254

## RESULT 4

X20248\_03  
 Continuation (4 of 10) of X20248 from base 300001 (Borrelia burgdorferi polynucleotide s  
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
 Fragment Name Begin End  
 WP X20248\_00 1 110000  
 WP X20248\_01 100001 210000  
 WP X20248\_02 200001 310000  
 WP X20248\_03 300001 410000  
 WP X20248\_04 400001 510000  
 WP X20248\_05 500001 610000  
 WP X20248\_06 600001 710000  
 WP X20248\_07 700001 810000  
 WP X20248\_08 800001 910000  
 WP X20248\_09 900001 910715

Query Match 2.2%; Score 37.8; DB 1; Length 110000;  
 Best Local Similarity 45.9%; Pred. No. 4.7;  
 Matches 129; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

OY 1379 TATATTCATATCAACAGACCATTCCTGGCATTCATTTAAATATATGGAAGTG 1438  
 DB 77062 TAAATTTAAAAAGAAAGAAATTTCAAAATCTCAAGCTCCATCATATTTCAAAATTA 77121

OY 1439 GATAGAGAACTGCAGCTCTCAATAGCGCTGATTTTGTGCAGTAATAATAA 1498  
 DB 77122 TATATTTAAAAATAGCAGTGAAGCTCAAAACCAAGGTTTTTCGAAACTTAAGTA 77181  
 OY 1499 AATCATTCATCCCTTTTGTGATTAATAATTTTCTAAATGATTTAGACTCTGTA 1558  
 DB 77182 TAGAGACTCCCTGCTCTCTTTGATTCATTTTAAATTAAGATTTAAACTTTTCA 77241  
 OY 1559 GGGGGGATATCTAATATATATAGTACATTTATACATAATGATTTCTGTAGGGGGG 1618  
 DB 77242 CTAATCTTTTCAAAATTAAGATAATTAATGAATCAATTCATTCATCAATCAATATC 77301  
 OY 1619 ATATTAATATGATTTTAGACTTCTGTAGGGGGGATTA 1659  
 DB 77302 TATTTAAATTAACATCAGAACTTCATATTTTGAAGATTA 77342

## RESULT 5

N90224  
 ID N90224 standard; DNA; 1052 BP.  
 AC N90224;  
 DT 1-NOV-1989 (first entry)  
 DE Malaria-specific DNA insert of clone 41-2.  
 KW DNA: malaria; clone 41-2; Plasmodium falciparum; protein; vaccine.  
 OS Plasmodium falciparum  
 PN EP-322712-A.  
 PD 05-JUL-1989.  
 PE 20-DEC-1988; 121299.  
 PR 30-DEC-1987; DE-831351.  
 PA (BEHM) Behringwerke.  
 PI Knapp B, Hundt E, Enders B, Kupper H;  
 DR WPI: 89-194071/27.  
 DR P-PDB: P90417.  
 PT New antigenic proteins from Plasmodium falciparum - new encoding nucleic  
 PT acid sequences and derived antibodies, useful in vaccines, diagnosis etc.  
 PS Claim 1; Table 17; 25pp; German.  
 CC The DNA is of a malaria-specific DNA insert of clone 41-2. The encoded  
 CC peptide (see P90417) is useful in a protective vaccine, esp. against  
 CC malaria. Produced antibodies are useful for passive immunisation, and  
 CC antibodies, the protein and the DNA sequence are useful as diagnostic  
 CC reagents. See also N90211-25 and P90403-19.  
 SQ Sequence 1052 BP: 394 A; 93 C; 88 G; 477 T;

Query Match 2.2%; Score 36.8; DB 1; Length 1052;  
 Best Local Similarity 50.0%; Pred. No. 0.96;  
 Matches 92; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

OY 1410 ATTACATTTTAAAAATATGTAAGAGTAGAGAGAACTGACGCTGCAATAGCCTAG 1469  
 DB 154 ATACAGTGTATGTAAGTGTCAAAAATTTATATGGAAGAAATTAATTAATATGTAA 213  
 OY 1470 GGCTGAATTTGTGCGATTAATAATAATCAATTCATCTTTTGTGATTAATAAT 1529  
 DB 214 TATATATATATATATATATATATATATTTCTTTCTTTCTTTTGTGTTT 273  
 OY 1530 TTTCAAAAATGATTTTACACTCTCTGAGGGGCGATATACATAATGATATAGTACT 1589  
 DB 274 ATATATGAAGTATATATATATCTTATGATGAGGATATATTAACAATATATGATTAAT 333  
 OY 1590 TTAT 1593  
 DB 334 AAT 337

## RESULT 6

V21205\_00/c  
 WP Sequence split into 17 fragments LOCUS V21205 Accession V21205  
 Fragment Name Begin End  
 WP V21205\_00 1 110000  
 WP V21205\_01 100001 210000  
 WP V21205\_02 200001 310000  
 WP V21205\_03 300001 410000

WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000  
WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000  
WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976  
ID V21209 standard; DNA: 1664976 BP.  
AC V21209;  
DE Methanococcus jannaschii circular chromosome.  
KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome; genome; autotrophic; extrachromosomal element; identification; ds.  
OS Methanococcus jannaschii.  
PN W09807830-A2.  
PD 26-FEB-1998.  
PE 22-AUG-1997; U14900.  
PF 22-AUG-1996; US-024428.  
PA (GENO-) INST GENOMIC RES.  
PA (UNIV ) UNIV ILLINOIS FOUND.  
PA (UJVO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;  
DR WPI: 98-169145/15.  
PR Complete genome sequence of methano-genic archaeon, Methanococcus jannaschii - useful in identification of M. jannaschii genome  
PT fragment  
PS Claim 13; Page 152-585; 614pp; English.  
CC The present sequence represents the complete 1.66-megabase pair genome sequence of the Methanococcus jannaschii circular chromosome. The present invention describes M. jannaschii open reading frames from the genome sequence. The invention also describes a computer based system for identifying fragments of the M. jannaschii genome that are homologous to target nucleotide sequences, comprising: (a) data storage means comprising the nucleotide sequence of the 1664976, 58407 or 16550 bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at least 99.9% identical to it; (b) search means for comparing a target sequence to the nucleotide sequence of the data storage means to identify a homologous sequence, and (c) retrieval means for obtaining the homologous sequence. The method, which is based on whole genome random sequencing of an autotrophic archaeon M. jannaschii, the genome of which consists of 3 physically distinct elements, a large circular chromosome (the 1664976 bp sequence given in V21209), a large circular extra-chromosomal element (the 58407 bp sequence given in V21210), and a small circular extra-chromosomal element (the 16550 bp sequence given in V21211), can be used in the identification of M. jannaschii genome  
CC fragment.  
CC Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T;  
SQ

Query Match 2.2%; Score 36.8; DB 1; Length 110000;  
Best Local Similarity 55.5%; Pred. No. 9;  
Matches 71; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 1427 TATGTGAGTGGATGAGGAGCTGCAATGAGCTTGGGCTGAATTTTGTGAC 1486  
DB 38254 TATATATGATTTATATATATGATTTATGATTTCTAGGAGTATTTTGAAGTATTTTNG 38195  
OY 1487 ATAAATAAATAAATCAATCATCTTTTGTGATTAATAATTTTCTAAATGTATTTT 1546  
DB 38194 AGACATFAAAAAATATATATATAATAATATATTTATTTATACCTTTCAAAATCTTTTAA 38135  
OY 1547 AGACTTCC 1554  
DB 38134 AAACATTC 38127

RESULT 2

V53366  
ID V53366 standard; DNA: 2854 BP.  
AC V53366;  
DE DNA encoding a Staphylococcus aureus protein of unknown function.  
KW Staphylococcus aureus protein; immune response induction; eye infection; antibody production; T-cell immune response; gastrointestinal infection; respiratory infection; inhibitor; bacterial infection; cardiac infection; central nervous system; kidney infection; urinary tract infection; antimicrobial compound identification; broad spectrum antibiotic; therapy; ss.  
KW Staphylococcus aureus.  
OS Ep-841394-A2.  
PN 13-MAY-1998.  
PD 13-MAY-1998.  
PE 24-SEP-1997; 307485.  
PF 24-SEP-1996; US-027032.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Black MT, Burnham MKR, Hodgson JF, Knowles DC,  
PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
PI Ward JM;  
DR WPI: 98-252940/23.  
DR P-PSDB: W75566.  
PT New nucleic acid sequences from Staphylococcus aureus MCHU29 - useful in vaccines and for treatment of bacterial infections of e.g. respiratory tract and central nervous system  
PS Claim 1; Page 69-70; 390pp; English.  
CC This sequence encodes a Staphylococcus aureus protein of unknown function, and represents a DNA sequence of the invention.  
CC The DNA sequences were isolated from Staphylococcus aureus MCHU29 (NCIMB 40771). Host cells containing the DNA sequences are used to produce polypeptides or fragments. The proteins are used in the treatment of disease, for inducing an immune response by administering them, to produce antibody and/or T-cell immune response. Antagonists of the CC proteins are used for the inhibition of bacterial polypeptides.  
CC Conditions which may be treated include bacterial infections, especially respiratory, cardiac, gastrointestinal, central nervous, eye, kidney, urinary tract, skin, bones and joints. The proteins can also be used to identify antimicrobial compounds which are broad spectrum antibiotics, especially useful in the treatment of H. pylori infection.  
SQ Sequence 2854 BP; 1012 A; 446 C; 453 G; 881 T;

Query Match 2.2%; Score 36.6; DB 1; Length 2854;  
Best Local Similarity 48.9%; Pred. No. 1.8;  
Matches 90; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

OY 1459 CAATGCGCTAGGGCGAATTTTGTGACATTAATAATTAATCATTCCTTTT 1518  
DB 1253 CAATGCGCGAGCGATTTATGACNACTATCAATTTTCTCANATGTGTATTTT 1322  
OY 1519 GATTATAAATTTTCTAAATGTATTTAGACTTCCTGTAGGGGCGATATACATAATGT 1578  
DB 1323 TATAANNNAAGGTAGAGCAAGGATTAATATATCTTCAATCGTTTCAATTTATATAT 1382  
OY 1579 ATATAGTACATTTATACATAATGTATTCCTGTAGGGGCGATATACATAATGTATTTAG 1638  
DB 1383 TTTTAAATTAAGGCGTAATGTTTAACTAAGCATATGACAGATGATATACAG 1442  
OY 1639 ACTT 1642  
DB 1443 AATT 1446

RESULT 8  
V74703  
ID V74703 standard; DNA: 2424 BP.  
AC V74703;  
DE 16-MAR-1999 (first entry)  
DE Staphylococcus aureus confg SEQ ID #392.  
KW Computer readable medium; vaccine; S.aureus infection; immunodetection; cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy; skin infection; surgical wound infection; scalded skin syndrome;









Mon Aug 9 13:26:10 1999

us-09-049-696-19.rng

Page 8

[illegible]

Search completed: August 6, 1999, 04:13:14  
Job time: 2146 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:19:32 ; Search time 3264.21 Seconds

(without alignments)  
1639.742 Million cell updates/sec

Title: US-09-049-696-19

Perfect score: 1683  
Sequence: 1 AACAAAGTGTGTCATCATC.....AAATGCTAAACAACCTGGGTA 1683

Scoring table: IDENTITY\_NWC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pac:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_se:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_vl:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hlg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hlg1:\*  
35: gb\_hlg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1671	99.3	3311	11	AF039400	Homo sapi

2	1528.4	90.8	2826	42	AF127036	AF127036 Homo sapi
3	942.6	56.0	2937	12	AB017156	AB017156 Mus muscu
4	790.8	49.0	878	5	195746	195746 Sequence 8
5	594.4	35.3	35278	11	AF039401	AF039401 Homo sapi
6	437.8	26.0	2984	3	BTG36445	BTG36445 Bos tauru
7	414.4	24.6	3317	3	AF001261	AF001261 Bos tauru
8	408.4	24.3	3022	12	AF047838	AF047838 Mus muscu
9	408.4	24.3	3471	12	AF052746	AF052746 Mus muscu
10	366.6	21.8	3415	42	AF043976	AF043976 Homo sapi
11	320.2	19.0	2820	3	AF001263	AF001263 Bos tauru
12	319.2	19.0	3288	3	AF001262	AF001262 Bos tauru
13	301.6	17.9	3604	9	AB026833	AB026833 Homo sapi
14	46.4	2.8	159453	34	AC006732	AC006732 Caenorhab
15	45.6	2.7	99083	36	CEY37A1B	CEY37A1B Caenorhab
16	43.8	2.6	256172	35	AC005139	AC005139 Plasmodiu
17	43.8	2.6	284972	35	AC005505	AC005505 Plasmodiu
18	43.2	2.6	112930	17	HSGEND	HSGEND Herpesvitu
19	43.2	2.6	43658	17	HSV3PRGEN	HSV3PRGEN Herpesvitu
20	43.2	2.6	110000	34	PFMAL4P1_0	PFMAL4P1_0 Plasmodiu
21	43.2	2.6	219200	35	AC006279	AC006279 Plasmodiu
22	43	2.6	91780	8	AC006192	AC006192 Arabidops
23	43	2.6	110000	34	PFMAL4P1_0	PFMAL4P1_0 Plasmodiu
24	42.4	2.5	164399	36	PFMAL3P6	PFMAL3P6 Plasmodiu
25	42.2	2.5	101679	8	ATF19B15	ATF19B15 Arabidops
26	42.2	2.5	97713	8	ATF25024	ATF25024 Arabidops
27	42	2.5	126356	34	AC006876	AC006876 Caenorhab
28	42	2.5	258033	35	AC005507	AC005507 Caenorhab
29	42	2.5	38174	37	CELE02H9	CELE02H9 Caenorhab
30	41.8	2.5	98067	9	HS435A7	HS435A7 Homo sapi
31	41.8	2.5	216293	11	HTCRBCHR9	HTCRBCHR9 Homo sapi
32	41.8	2.5	4601	36	DMU11584	DMU11584 Drosophila
33	41.8	2.5	19517	37	DMU37541	DMU37541 Drosophila
34	41.6	2.5	1231	1	MC082	MC082 M. capricolu
35	41.6	2.5	119704	7	OSCHLPXX	OSCHLPXX O. sinensis
36	41.6	2.5	228625	35	AC004153	AC004153 Plasmodiu
37	41.6	2.5	300172	35	AC005308	AC005308 Plasmodiu
38	41.6	2.5	220475	35	AC006278	AC006278 Plasmodiu
39	41.6	2.5	231699	35	AC006281	AC006281 Plasmodiu
40	41.4	2.5	7218	5	166494	166494 Sequence 14
41	41.4	2.5	53778	8	TL11	TL11 AF183933 Arabidops
42	41.4	2.5	154902	11	AC005799	AC005799 Homo sapi
43	41.4	2.5	199606	35	AC004688	AC004688 Plasmodiu
44	41.4	2.5	196149	35	AC004709	AC004709 Plasmodiu
45	41.4	2.5	75551	37	AC005111	AC005111 Drosophila

## ALIGNMENTS

RESULT 1

AF039400 3311 bp mRNA PRI 15-DEC-1998

LOCUS Homo sapiens calcium-dependent chloride channel-1 (hcccl1) mRNA, complete cds.

ACCESSION AF039400

NID 94009457

VERSION AF039400.1 GI:4009457

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 3311)

AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.

TITLE Genomic cloning, molecular characterization, and functional analysis of human hcccl1, the first human member of the family of Ca2+-activated Cl- channel proteins

JOURNAL Genomics 54 (2), 200-214 (1998)

MEDLINE 99047526

REFERENCE 2 (bases 1 to 3311)

AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.

TITLE Direct Submission

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

## FEATURES

source

Location/Qualifiers

1. 3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. 3311  
/gene="hclcal1"  
352..3096  
/gene="hclcal1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="pid:g4009458"  
/db\_xref="gi:4009458"  
/translation="WGPFRSSVFIIILHLEGLASNSLIQLNNNGEIVVAIDPNP  
EDETLIQIKDMVTOASLYLEATGKRFYRNVALILPETWTKADYVRLLEYKNA  
DIVLAESTPPDENDEYTPDMNGNCEKGRILHTPELPIGKLAEGPQKAFVEMAH  
LRMGVFEYNDNDEKPYLSNGRIQAVRGAIGTGVVKKCGSGCYTKRCPFNVTGL  
YEKGEFYLQSRQREKASIMRAQHDSTVECTEONHKRPAKNOKCNIRSTREVI  
RDSDFKRTTPHTTOPPPNPTSLDQIGRTYCLVDKSGSATGKRLKRLNAGVFL  
LQTVELGSMVAVTFDSAHVQSELIQINSVDKDTLAKRLPAASGGTSLCSGLRSA  
FTVIRKKYPTGSEIVLITDEDNITSGCFNEVKOSGAIHTVLAQPSMAQAGVFL  
MTGLOIYASDOVONNGLIDAFGLASNGAVSQRSLQLESGLTIQNSOMNGVIY  
DSVKGKDLFLITTPTOPLIQIMDPSGKOGFVDYDKTKMAVLIQIGIKGVWKY  
SLOASQTLTLTVSRASNATLPTVTSKNTKSRKPSPLVYVYANRQASPLTRA  
SVTALIESVNKRTVLELLDNGAGADATKDDGVISRYTITDNGRISYKRALGAVN  
AARRVIVQSGALIPGWIENDELQMPNPEIKDVQHKQVCFSTSVKSGSLGASV  
DVNAPRIDLEPPQGITDLKAEIHGSLINLTWAPGDYDGHAKYIIRISISLD  
LRDFNESIQVNTALIPKEANSEVFLFKENITFENGDTLFLIAQVDRVDKSEI  
SNIRVSLFIPQTPPETPSPDETSAFCPNHINSTIPGILHILKIMKIGELQISIA  
"

CDS

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 99.3%; Score 1671; DB 11; Length 3311;  
Best Local Similarity 99.9%; Pred. No. 0;

Matches 1682; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 AACAAAGTGGCCATCATCACACAGTGGCTTGGGGCCCTCTGACGCTCAAGAACTAG 60  
DB 1628 AACAAAGTGGCCATCATCACACAGTGGCTTGGGGCCCTCTGACGCTCAAGAACTAG 1687  
OY 61 AGGAGCTGCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGANA 120  
DB 1688 AGGAGCTGCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGANA 1747  
OY 121 ATGGCCTATTGATGCTTTGGGCCCTTTCATCAGGAATGGAGCTGCTCTCAGCGCT 180  
DB 1748 ATGGCCTATTGATGCTTTGGGCCCTTTCATCAGGAATGGAGCTGCTCTCAGCGCT 1807  
OY 181 CCATCCAGCTGAGAGTAAGGATTAACTCCAGAACAGCCAGTGGATGAATGGCAGC 240  
DB 1808 CCATCCAGCTGAGAGTAAGGATTAACTCCAGAACAGCCAGTGGATGAATGGCAGC 1867  
OY 241 TGATGCTGACGACGACCGTGGGAAGGACATTTGTTTTCACCTGGAGAAAGCAGC 300  
DB 1868 TGATGCTGACGACGACCGTGGGAAGGACATTTGTTTTCACCTGGAGAAAGCAGC 1927  
OY 301 CTCGCCAAATCTTCTCTGGATGCCAGTGGACAGAGCAAGAGTGGCTTTGATGGACA 360  
DB 1928 CTCGCCAAATCTTCTCTGGATGCCAGTGGACAGAGCAAGAGTGGCTTTGATGGACA 1987  
OY 361 AAAACACCAAAATGGCTCACTCAAAATCCAGGCAATGCTAAAGTTGGCACTTGGAAAT 420  
DB 1988 AAAACACCAAAATGGCTCACTCAAAATCCAGGCAATGCTAAAGTTGGCACTTGGAAAT 2047  
OY 421 ACAGTCTGAGCAAGCTCACAAACCTTGACCTGACGTCCCTGCTGCTCAATG 480

DB 2048 ACAGTCTGAGCAAGCTCACAAACCTTGACCTGACGTGACGCTCCGCTGCTCAATG 2107  
OY 481 CTACCCCTGCTCCAAATTCACAGGACTTCCAAAAGCAAGACACGCAAAATTCGCCA 540  
DB 2108 CTACCCCTGCTCCAAATTCACAGGACTTCCAAAAGCAAGACACGCAAAATTCGCCA 2167  
OY 541 GCCCTCTGTAAGTTATTCGCAAAATATTCGCCAAGGACCTCCCAATTCACAGGCGCAGTG 600  
DB 2168 GCCCTCTGTAAGTTATTCGCAAAATATTCGCCAAGGACGCTCCCAATTCACAGGCGCAGTG 2227  
OY 601 TCACAGCCCTGATTAATCAGTGAATGGAATAACAGTACCTTGGAACTACTGCTAATG 660  
DB 2228 TCACAGCCCTGATTAATCAGTGAATGGAATAACAGTACCTTGGAACTACTGCTAATG 2287  
OY 661 GAGCAGGTGCTGATCTCTAAGATGAGTGAGGCTGTCTACTCAAGATTTTCAAACTTATG 720  
DB 2288 GAGCAGGTGCTGATCTCTAAGATGAGTGAGGCTGTCTACTCAAGATTTTCAAACTTATG 2347  
OY 721 ACACGAATGTAGATACAGTGTAAAGTGGCGGCTCTGGAGAGTTAAACGACGACAGC 780  
DB 2348 ACACGAATGTAGATACAGTGTAAAGTGGCGGCTCTGGAGAGTTAAACGACGACAGC 2407  
OY 781 GGAAGTATATACCCACGACAGAGTGGACACTGTACATACCTGGCTGGATTGAGAAATATG 840  
DB 2408 GGAAGTATATACCCACGACAGAGTGGACACTGTACATACCTGGCTGGATTGAGAAATATG 2467  
OY 841 AAATCAATGGAATCCACCAAGACCTGAAATTAATTAAGATGATTTAAACAAAGCAAG 900  
DB 2468 AAATCAATGGAATCCACCAAGACCTGAAATTAATTAAGATGATTTAAACAAAGCAAG 2527  
OY 901 TGTGTTTACAGAAACATCTCTGGGAGGCTCATTTGGCTTGTGATGCCAAATGCTC 960  
DB 2528 TGTGTTTACAGAAACATCTCTGGGAGGCTCATTTGGCTTGTGATGCCAAATGCTC 2587  
OY 961 CCATACCTGATCTCTCTCCACCTGGCCAAATCACCGACCTGAAGCGGAATTCACGGGG 1020  
DB 2588 CCATACCTGATCTCTCTCCACCTGGCCAAATCACCGACCTGAAGCGGAATTCACGGGG 2647  
OY 1021 GCAGTCTATTAATCTGACTTGGACAGCTCCTGGGATGATTAATGACCATGGAACAGCTC 1080  
DB 2648 GCAGTCTATTAATCTGACTTGGACAGCTCCTGGGATGATTAATGACCATGGAACAGCTC 2707  
OY 1081 ACAAGTATATCATTTGGAATTAAGACATGATTTGATCTGACAGACAGTTCAATGAT 1140  
DB 2708 ACAAGTATATCATTTGGAATTAAGACATGATTTGATCTGACAGACAGTTCAATGAT 2767  
OY 1141 CTTCTCAAGTGAATACCTGCTGTCATCCCAAGGAAGCAACCTGAGGAAGTCTTT 1200  
DB 2768 CTTCTCAAGTGAATACCTGCTGTCATCCCAAGGAAGCAACCTGAGGAAGTCTTT 2827  
OY 1201 TGTTTAAACAGAAAACATTAATTTTGAATATGACAGATCTTTTCAATTCACAG 1260  
DB 2828 TGTTTAAACAGAAAACATTAATTTTGAATATGACAGATCTTTTCAATTCACAG 2887  
OY 1261 CTGTTGATTAAGTGTGATGAAATACAGAAATATCCAAATTCGACAGATATCTTTGTTA 1320  
DB 2888 CTGTTGATTAAGTGTGATGAAATACAGAAATATCCAAATTCGACAGATATCTTTGTTA 2947  
OY 1321 TTCTCCACAGACTCCGGCACAAGACACTAGTCCGTGAAGAAAGCTGCTCTGTCTTA 1380  
DB 2948 TTCTCCACAGACTCCGGCACAAGACACTAGTCCGTGAAGAAAGCTGCTCTGTCTTA 3007  
OY 1381 ATATTCAATACAGACACCAATTCCTGGCATTCACATTTTAAATTAATGAGAAAGTGA 1440  
DB 3008 ATATTCAATACAGACACCAATTCCTGGCATTCACATTTTAAATTAATGAGAAAGTGA 3067  
OY 1441 TAGGAGAACTGACAGCTGTCAATAGCTTAGGCTGAATTTTGTGACAGATAAATAAATAA 1500  
DB 3068 TAGGAGAACTGACAGCTGTCAATAGCTTAGGCTGAATTTTGTGACAGATAAATAAATAA 3127  
OY 1501 TCATTCAATCTTTTGTGATTAATAATTTTCAAAATGATATTTGACATCTGCTAGG 1560

Db 3128 TCATCATC-TTTTTGTATATAAATTTCTATAATGATTAGACTCTCTAGG 3186  
QY 1561 GGGCGATATCTAATAGTATAGTACATTATCTAATAGTATCTCTAGGCGCAT 1620  
Db 3187 GGGCGATATCTAATAGTATAGTATAGTATCTAATAGTATCTCTAGGCGCAT 3246  
QY 1621 ATACTAATGATTATTTAGACTCTCTAGGCGCATATAAATGCTAAACAACCTG 1680  
Db 3247 ATACTAATGATTATTTAGACTCTCTAGGCGCATATAAATGCTAAACAACCTG 3306  
QY 1681 GTA 1683  
Db 3307 GTA 3309  
RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 2826)  
AUTHORS Agnel, M., Vernal, T. and Culouscou, J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel, M. and Culouscou, J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Ruell-Malmaison 92500, France  
FEATURES  
source  
1..2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue-type="small intestine; colon"  
1..2826  
/gene="CACCL"  
5..2749  
/gene="CACCL"  
/note="bovine epithelial chloride channel homolog"  
/product="calcium-activated chloride channel protein 1"  
/codon\_start=1  
/protein\_id="AAD23487.1"  
/db\_xref="PIR:4585469"  
/db\_xref="GI:4585469"  
/translation="MGPFSVFIILHLHLEGLASNSLIQLNNNGEGLVAVIDPNV  
EDETLIQIKMVAQASLYIFATIGKREFKFAVILIPETTKADYAPKLETKNA  
DYVAESTPRGNDERTBOMGCGEKEHILTPPIAKKALAEIGPOGRATYHEBAH  
YKGEFVYQKQTERKASIMFAQVDSIYEICTEDHNEAPNOKNCKENSTVEI  
LROSEDKTTPMTOPNPPTFSLQIGRIYCLVLDGSGMAIGNLNLNAGOLEFL  
LOTVELGSMWVTFDSAAHVOSELIQINSGRDRLAKRLPAASGGSTGSGRSA  
FTVIRKPTDSEIVLLTDEDNITSGCFENYKOSGALIHVLAQPSAOLHELESK  
MTGIGQYASDVOYNNGLIDAGALSNGGAVSONSIQESGLLQNSQKMGATVIV  
DSTVGDHLELITWTQPPQILLPDSGGKGGFVADKTKAAVLIQPIAVGVTKY  
SLQASQTLTIVTSRASNATLPITVYSKTKNDTSKPSPLVYVANIQOGAPILRA  
SVTALIESVNGKTVLELDNGAGADATDDGVYSKPTTYDTNGSYSVKVALGVA  
AARRVYPOOGSALYIGWIENDEIOMNPREINKDDVOHOCFSRSTSGSVAS  
DYVNAIPDLPFGQITDKAKIHGSLNLTAKGDYDHDGTAHKYIRISTID  
LQDKNESLQVNTTALIRKANSSEYFLERKPNITFENKTDLFIAIQAVDKYDLSSEI  
SNIAVSLFIPQTPETPSPDSTSPCPNHNINSTIPGILHLIKIMKWIIGLQLSIA

Query Match 90.8%; Score 1528.4; DB 42; Length 2826;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1532; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 AACAAAGTGGTGCATCATCATCACAGTGGCTTTGGGGCCCTGGAGCTCAAGAACTAG 60  
Db 1281 AACAAAGTGGTGCATCATCATCACAGTGGCTTTGGGGCCCTGGAGCTCAAGAACTAG 1340  
QY 61 AGGAGCTGTCCAAATAGCAGAGGATTACAGACATATGCTTCAGATCAAGTTCAGACACA 120  
Db 1341 AGGAGCTGTCCAAATAGCAGAGGATTACAGACATATGCTTCAGATCAAGTTCAGACACA 1400  
QY 121 ATGGCTCATGTATGCTTTTGGGCCCCCTTTCATCAGGAAATGAGCTGTCTCAGCCCT 180  
Db 1401 ATGGCTCATGTATGCTTTTGGGCCCCCTTTCATCAGGAAATGAGCTGTCTCAGCCCT 1460  
QY 181 CCATCCAGCTTGAGTAAGGGAATTAAACCTCCAGAACAGCCAGTGGATGAATGGACAG 240  
Db 1461 CCATCCAGCTTGAGTAAGGGAATTAAACCTCCAGAACAGCCAGTGGATGAATGGACAG 1520  
QY 241 TGATCGTGCACAGCACCGTGGGAAAGAGACATTGTTCTTATCACCTGGACACAGCAGC 300  
Db 1521 TGATCGTGCACAGCACCGTGGGAAAGAGACATTGTTCTTATCACCTGGACACAGCAGC 1580  
QY 301 CTCGCCAATCTCTCTCTGGGATCCAGTGCAGACAGACAGAGTGGCTTTGTAGTGACA 360  
Db 1581 CTCGCCAATCTCTCTCTGGGATCCAGTGCAGACAGACAGAGTGGCTTTGTAGTGACA 1640  
QY 361 AAAAACAACAAAAGGCTCTCAATCCCAATCCAGGCACTTGTAGTGGACACTGGAAAT 420  
Db 1641 AAAAACAACAAAAGGCTCTCAATCCCAATCCAGGCACTTGTAGTGGACACTGGAAAT 1700  
QY 421 ACAGTGCAGACAGCTCAACAACTTGACCTGACGTGACGTCGCTCCCAATG 480  
Db 1701 ACAGTGCAGACAGCTCAACAACTTGACCTGACGTGACGTCGCTCCCAATG 1760  
QY 481 CTACCTGCTGCTCAATTAAGTACTTCCAAACGAAACAGACACAGCAATTTCCCA 540  
Db 1761 CTACCTGCTGCTCAATTAAGTACTTCCAAACGAAACAGACACAGCAATTTCCCA 1820  
QY 541 GCCCTGTGATGTTATGCAAAATATTCGCAAGAGACCTCCCAATTTCCAGGGCCAGTG 600  
Db 1821 GCCCTGTGATGTTATGCAAAATATTCGCAAGAGACCTCCCAATTTCCAGGGCCAGTG 1880  
QY 601 TCACAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660  
Db 1881 TCACAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1940  
QY 661 GAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720  
Db 1941 GAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2000  
QY 721 ACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780  
Db 2001 ACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2060  
QY 781 GAGAGTGTATCCCGACAGAGTGAAGCACTGTATACCTGGCTGGATGGAATGATG 840  
Db 2061 GAGAGTGTATCCCGACAGAGTGAAGCACTGTATACCTGGCTGGATGGAATGATG 2120  
QY 841 AATATACATGGAATCCACAGACCTGGAATTAATAGATGATGTTCAACACAGCAAG 900  
Db 2121 AATATACATGGAATCCACAGACCTGGAATTAATAGATGATGTTCAACACAGCAAG 2180  
QY 901 TGGTGTTCAGCAAGATCTCGGGAGGCTCATTTGGCTCTGATGTCACCAATGCTC 960  
Db 2181 TGGTGTTCAGCAAGATCTCGGGAGGCTCATTTGGCTCTGATGTCACCAATGCTC 2240  
QY 961 CCATACCTGATCTCTTCCACCTGGCCCAATACCGCACTGGAAGGGGAATTCACGGGG 1020  
Db 2241 CCATACCTGATCTCTTCCACCTGGCCCAATACCGCACTGGAAGGGGAATTCACGGGG 2300



Oy	661	GAGCAGGCGCTGATGCTACTTAAGAGAGCGCTCTACACGAATTTTCAACATTTG	720
Oy	1954	GAGCAGGCTGCGATGCGCCACCAAGATATATGTCTTACTCAAGTTTTCACAGCTTTTG	2013
Oy	721	ACAGCAATGCTGATGTATACAGTGTAAAGTGGGGCTCTGGGAGGAGTTTAAACGACCCAGAC	780
Db	2014	ATGCAATATGATAGATATACAGGTTTAAATATATGGCTCTGGGAGGAGTCACTTACAGCAAC	2073
Oy	781	GGAAGCTGATATCCCACAGAGTGGAGCACTGTACATACCTGGCTGGATTGGAATGATG	840
Db	2074	AGAGAGCGACCTCCGAGAGACAGACCCTATGTACATTAAGTGGCTGGATTGAGGATGGTG	2133
Oy	841	AAATACATGGAATTCACCAAGACCTGAATTTAATAGATGATGTGTCAACCAAGCAAG	900
Db	2134	AAGTACAGATATACCCCAACCAAGCTCTGAACCTA-----GTTATGTTCAAGCAAGCAAGC	2187
Oy	901	TGTGTTTCAGCAGAACATCTCTGGGAGGCTCATATTGTGGCTTCTGATGTGC---CNAATG	957
Db	2188	TGTGTTTCAGCAGAACATCTCTGGGAGGCTCATATTGTGGCTTCTGATGTGC---CNAATG	2247
Oy	958	CTCCCATTAACCTGATCTCTTCCCACTGGCCAAATTCACCACTGAAAGCGGAAATTCAG	1017
Db	2248	CTCCCATTAACCTGATCTCTTCCCACTGGCCAAATTCACCACTGAAAGCGGAAATTCAG	2307
Oy	1018	GGGGAGCTCATTAATCTGACTTGGACACTCTGAGGAGTGAATATGACATGGGAAG	1077
Db	2308	GGCAGAACCTGGTATATCTGACTTGGACACTCTGAGGAGTGAATATGACATGGGAAG	2367
Oy	1078	CTCACAGATATATCATTCGATAGTACAAAGTATTCGTATCTCAGAGCAAGTTCATG	1137
Db	2368	CTTCCAACTATCATTCATTCGATAGTACAAAGTATTCGTATCTCAGAGCAAGTTCATG	2427
Oy	1138	AATCTCTTCAAGTATATACACTACTGCTCTATCCCAAGAGAACCACTTGAGAACTCT	1197
Db	2428	CCTCACTTCCAAAGTATATACACTACTGCTCTATCCCAAGAGAACCACTTGAGAACTCT	2487
Oy	1198	TTTTGTTTAAACCGAAGAAATTAATCACTTTGAAATGGACAGATCTTTTCAATGGCTATTC	1257
Db	2488	TTTGATTTGAATCTGGAGGCAACACTTTTGGAAATGGACAGATCTTTTCAATGGCTATTC	2547
Oy	1258	AGGCTGTGATAGTGTGATCTGAAATCAGAAATATCAACAATTCAGACAGATATTTGT	1317
Db	2548	AGGCTGTGATAGTGTGATCTGAAATCAGAAATATCAACAATTCAGACAGATATTTGT	2607
Oy	1318	TATATCTCCACAGACTCTCGCCGACAGACACTAGTCTGATGAACGTCTGCTCTTGTCTC	1377
Db	2608	TATATCTCCACAGACTCTCGCCGACAGACACTAGTCTGATGAACGTCTGCTCTTGTCTC	2658
Oy	1378	CTAATATTCATATCAACAGCACCATTCTGGGACTTCAATTTTAAAAATTAATGSGAAGT	1437
Db	2659	CTGACATTCAGCATATCAACAGCACCATTCTGGGACTTCAATTTTAAAAATTAATGSGAAGT	2718
Oy	1438	GGATAGGAGAACTGCACTCTCAATAGCCTAGGCGCTGAATTTTGTCAAGATTAATTAAT	1497
Db	2719	GGATAGGAGAACTGCACTCTCAATAGCCTAGGCGCTGAATTTTGTCAAGATTAATTAAT	2778
Oy	1498	AAATCATTTCACTCTTTTGTGATATATAAATTTTCTAAAAATGATTTTAAAGCTTCTGT	1557
Db	2779	A-----GTCATCTCTCTTCACTGAGAAATTTCTTAAAAATGATCTTAGACTCTGT	2830
Oy	1558	AGGGGGCGATATACTAA 1574	
Db	2831	AGGGGGCGGATATAGTA 2847	
RESULT	4		
LOCUS	195746	878 bp	DNA
DEFINITION	Sequence 8 from patent US 5733748.		PAT
ACCESSION	195746		
VERSION	g3940216		
	195746.1	GI:3940216	
			17-JUL-1998

KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 878)
TITLE	Yu, G. and Rosen, C.
JOURNAL	Col. on specific genes and proteins
FEATURES	Patent: US 5733748-A 8 31-MAR-1998;
source	Location/Qualifiers
	1..878
	/organism="unknown"
BASE COUNT	257 a 179 c 188 g 241 t 13 others
ORIGIN	
Query Match	47.0% Score 790.8; DB 5; Length 878;
Best Local Similarity	97.9% Pred. No. 3.4e-181;
Matches 820; Conservative	1; Mismatches 15; Indels 2; Gaps 2
QY	692 TGTCTACTCAAGGATATTTTCACAACTTATGACAGATGGTAGATACAGTAAAGTGGC 751
DB	1 TGTCTACTCAAGGATATTTTCACAACTTATGACAGATGGTAGATACAGTAAAGTGGC 60
QY	752 GCGCTGGGAGGAGTTAACGCGACGACGAGAGTATACCCACGACAGTGGAGCACT 811
DB	61 GCGCTGGGAGGAGTTAACGCGACGACGAGAGTATACCCACGACAGTGGAGCACT 120
QY	812 GTACATACCTGGCTGGATTGAGATGATGAATACATGGAATCCACCAAGCCTGGAAT 871
DB	121 GTACATACCTGGCTGGATTGAGATGATGAATACATGGAATCCACCAAGCCTGGAAT 180
QY	872 TAATAAGATGATGTTCAACACACAGCAAGTGTGTTTCAGCAGAACATCCTCGGAGGCTC 931
DB	181 TAATAAGATGATGTTCAACACACAGCAAGTGTGTTTCAGCAGAACATCCTCGGAGGCTC 240
QY	932 ATTGTGGCTTGTGATGTGCCAAATGCTCCCATCTGATCTCTTCCACCTGGCCAAAT 991
DB	241 ATTGTGGCTTGTGATGTGCCAAATGCTCCCATCTGATCTCTTCCACCTGGCCAAAT 300
QY	992 CACGACCTGGAAGGGGAAATCACGGGGGCGAGTCTCATTAATCTGACTTGGACAGCTCC 1051
DB	301 CACGACCTGGAAGGGGAAATCACGGGGGCGAGTCTCATTAATCTGACTTGGACAGCTCC 360
QY	1052 TGGGATGATTTATGACCATGGAACAGCTCACAGTATATCATTTGCAATATGATCAAGTAT 1111
DB	361 TGGGATGATTTATGACCATGGAACAGCTCACAGTATATCATTTGCAATATGATCAAGTAT 420
QY	1112 TCTTATCTCGAGAGCAAGTTCATGAAATCTCTTAAGGAATTAATAGCTGCTCATCCC 1171
DB	421 TCTTATCTCGAGAGCAAGTTCATGAAATCTCTTAAGGAATTAATAGCTGCTCATCCC 480
QY	1172 AAAGGAGCGCAACTGTGAGGAGTCTTTTGTGTTTAAACGAGAAACATTACTTTGAAAA 1231
DB	481 AAAGGAGCGCAACTGTGAGGAGTCTTTTGTGTTTAAACGAGAAACATTACTTTGAAAA 540
QY	1232 TGGCACAAGATCTTTTCATTTGCTATTCAGGCTGTATATAAGTTCATCTGAATTCGAAT 1291
DB	541 TGGCACAAGATCTTTTCATTTGCTATTCAGGCTGTATATAAGTTCATCTGAATTCGAAT 600
QY	1292 ATCCAAATTTGACAGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAGACACACTAG 1351
DB	601 ATCCAAATTTGACAGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAGACACACTAG 660
QY	1352 TCTTGATGAAACGCTGCTCTGT-T-CCTAATATTCATATCAACAGACCAATCTCTGGCA 1410
DB	661 TCTTGATGAAACGCTGCTCTGTCTTGTCTTAATATTCATATCAACAGACCAATCTCTGGCA 720
QY	1411 TTCACATTTTAAAAATTATGTGGAAGTGGATAGAGAACTGAGCTGTCATATAGCTAGG 1470
DB	721 TTCACATTTTAAAAATTATGTGGAAGTGGATAGAGAACTGAGCTGTCATATAGCTAGG 780
QY	1471 GCTGAATTTTGTGCAGATAAATTAATTAATCATCATCTTTTGTGATTTAAAA 1528

Db 781 GGTGAATTTTCTGGCGTGAAT-AAATAATSAATTTCANCCITTTTGTGTTATATAA 837

RESULT 5  
AF039401  
LOCUS AF039401 35278 bp DNA PRI 15-DEC-1998  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.  
ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human CLCA1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="L1 transposable element"  
/rpt\_type-dispersed  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..31840,32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PIR:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPFKSSVFLIIHLHLEGALSNSLIOLNNNGYRCIYVADPNP  
EDFLIQIKDNYTOASLYLFEATGKRFYKRNVAII,PERWKTKADYVRKLETKNA  
DVAIAESTIPGNDPEYTEOMGNGEGERIHLTPDFLAGKLAEGPQGAFAEYNAH  
LRNGVDEYNNDEKFLYSNGRIQAVRCSAGITGTINVKKCGSGSCYTRKCTFKVIGL  
YEGGCEFLVLOSROTAKASIMFAQHDVDSIVFECTEONHNKRAPNKKONKCLRSTVEYI  
RSEDEFKKTTPMTQPNPTEFSLIIGQRIYCLVDRSGSMATGNRLNRLNQAQGLFL

LQTVELGSWVGWYTFPSAAHVQSELIQINGSRDILAKRLPAASGRTSCGLNSA  
FIVIRKRPIDGSEIVLIDGEDNTISGCNEVKQSAIHTVALPSSAQDELELSK  
MTGGLQTYASDOVONNGLIDAFGLSSNGAVSQRSIOLESKGLTIONSQMNQVYIY  
DSTVQKDTLELITWTQPPQILIMDPSGKQGFVVDKNTKAYLQIPGIAKGVTKWY  
SLQASQDTLELITWTQPPQILIMDPSGKQGFVVDKNTKAYLQIPGIAKGVTKWY  
SVTALIESVQSGALYVTELELDNGAGADATKDDGYSRFTYDNGRYSVRALGVN  
AARRVYRQSGALYTPGTEENDELOMNPRPEINNKDDQHKOVCSRTSSGSPVAS  
DVPNAPIDLPFGQITIDLNAEIHGSLINLWTAFGDDIDHGTAKYIIRISTSLD  
LRKFNESLQVNTALIPKEANSEVFLFKPENITFENGTDLEIAQAVDRVLDKSEI  
SNIRARVSLFIPQIPETPSDETSAPCPNHIINSTIPGIHLIKIMKWIGELQLSIA  
"

6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15

BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN

Query Match 35 3%; Score 594.4; DB 11: Length 35278;  
Best Local Similarity 99.7%; Pred. No. 9.4e-134;  
Matches 606; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1076 AGCTCACAAGATATATCATTCGGAATTAAGTACAGATATCTTGATCTCAGAGACAAGTTCAA 1135  
|||||  
Db 32917 AGCTCACAAGATATATCATTCGGAATTAAGTACAGATATCTTGATCTCAGAGACAAGTTCAA 32976

QY 1136 TGAATCTCTTCAAGTGAATACACTACTGCTCTCATCCCAAGGAAGCAACTCTGAGAGAGT 1195  
|||||  
Db 32977 TGAATCTCTTCAAGTGAATACACTACTGCTCTCATCCCAAGGAAGCAACTCTGAGAGAGT 33036

QY 1196 CTTTGTGTTAAACCAAAACAACTTACTTTGAAAAATGCACAGATCTTTCATTGCTAT 1255  
|||||  
Db 33037 CTTTGTGTTAAACCAAAACAACTTACTTTGAAAAATGCACAGATCTTTCATTGCTAT 33096

QY 1256 TCAAGGCTGTTGATTAAGTCCATCTGAAATCAGAAATATCCAATTCAGACAGATATCTTT 1315  
|||||  
Db 33097 TCAAGGCTGTTGATTAAGTCCATCTGAAATCAGAAATATCCAATTCAGACAGATATCTTT 33156



QY 1316 GTTATTCCTCCACAGACTCCGCCAGACACTAGTCTGATGAAACGTCTCTCTTG 1375  
|||||  
DB 33157 GTTATTCCTCCACAGACTCCGCCAGACACTAGTCTGATGAAACGTCTCTCTTG 33216  
QY 1376 TCCATATTCATATCAACAGCAGCACTCTGCGATTCACATTTAAATTAATGAGAA 1435  
|||||  
DB 33217 TCCATATTCATATCAACAGCAGCACTCTGCGATTCACATTTAAATTAATGAGAA 33276  
QY 1436 GTGATAGAGAGAACTCAGCTGTCAATAGCCTAGGCGTGAATTTTGTCCAGATAATAAA 1495  
|||||  
DB 33277 GTGATAGAGAGAACTCAGCTGTCAATAGCCTAGGCGTGAATTTTGTCCAGATAATAAA 33336  
QY 1496 ATAAATCATTCATCTTTTGTGATTAATAATTTTCTAAATGTAATTTTAACTTCT 1555  
|||||  
DB 33337 ATAAATCATTCATCTTTTGTGATTAATAATTTTCTAAATGTAATTTTAACTTCT 33395  
QY 1556 GTAGGGGGGATTAATGTAATAGTATTAATTAATTAATGTAATGTAATGTAATGTA 1615  
|||||  
DB 33396 GTAGGGGGGATTAATGTAATAGTATTAATTAATTAATGTAATGTAATGTAATGTA 33455  
QY 1616 GCGATTAATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 1675  
|||||  
DB 33456 GCGATTAATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 33515  
QY 1676 ACTGGGTA 1683  
|||||  
DB 33516 ACTGGGTA 33523  
RESULT 6  
BTU36445 2984 bp mRNA MAM 09-FEB-1996  
LOCUS BTU36445 2984 bp mRNA MAM 09-FEB-1996  
DEFINITION Bos taurus calcium-activated chloride channel mRNA, complete cds.  
ACCESSION U36445  
NID 9184065  
VERSION U36445.1 GI:1184065  
KEYWORDS  
SOURCE  
ORGANISM  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE  
1 (bases 1 to 2984)  
Cunningham, S.A., Awayda, M.S., Bublief, J.K., Ismailov, I.I.,  
Arrate, M.P., Berdiev, B.R., Benos, D.J., and Fuller, C.M.  
Cloning of an epithelial chloride channel from bovine trachea  
J. Biol. Chem. 270 (52), 31016-31026 (1995)  
JOURNAL 96125078  
MEDLINE 96125078  
REFERENCE 2 (bases 1 to 2984)  
Benos, D.J.  
Direct Submission  
Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,  
University of Alabama at Birmingham, BHSB 706, Birmingham, AL  
35294, USA  
FEATURES  
source  
Location/Qualifiers  
1..2984  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19..2730  
/note="epithelial chloride channel"  
/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="p3d:9184066"  
/db\_xref="gi:1184066"  
/translation="VPRVLTFLTLHLPGKSSMVLNNGYDGLVIAINPSVE  
DEKLIONKEMTEASTYLEHATKRYFRNYSIIPTWKSSEYLPKQSDQAE  
VIVANPYLKHGDDPYTIOYGRGEGGYHFPNPLTNLPIYSGRGAFAVEVAHL  
RMGIDENYDGPYISRNTEATRCSTHIGTIVYACOGSCITRRCRDSOTGL  
YEAKCTFIPEKQTARESIMFQSLHSTEFCTETHNVEANLONKRCNGKSTMDVI  
CDS  
1..2984  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19..2730  
/note="epithelial chloride channel"  
/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="p3d:9184066"  
/db\_xref="gi:1184066"

Query Match 26.0%; Score 437.8; DB 3; Length 2984;  
Best local similarity 60.2%; Pred. No. 6.4e-96;  
Matches 809; Conservative 0; Mismatches 507; Indels 27; Gaps 4;  
BASE COUNT 1027 a 583 c 561 g 813 t  
ORIGIN  
1 ACACAAAGTGGTCCATCATCACACAGTGGTGGGCGCTCTGAGCTCAAGAACTAG 60  
|||||  
DB 1304 AACAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1363  
QY 61 AGAAGCTGCCAAATAGACAGAGGTTTACAGCATATGCTTCAGATCAAGTTCAAGACA 120  
|||||  
DB 1364 AGAAGCTGCCAAATAGACAGAGGTTTACAGCATATGCTTCAGATCAAGTTCAAGACA 1417  
QY 121 ATGGCTCATTTGATGCTTTGGGCGCTTTTCATCAGAAATGAGCTGTCTCAGCGCT 180  
|||||  
DB 1418 ATGGCTCATTTGATGCTTTGGGCGCTTTTCATCAGAAATGAGCTGTCTCAGCGCT 1477  
QY 181 CCATCCAGCTTGAGAGTAAGGATTAACCTCCAGAACGCCAGTGGATGACACAG 240  
|||||  
DB 1478 CTAATGAGTTGGAAAGCAAGCTTGGCAATTAACAAAGAAATGTAATGTAACGG 1537  
QY 241 TGAATGAG 300  
|||||  
DB 1538 TGAATGAG 1597  
QY 301 CTTCCCAAAATCTTCTCTGAGATCCAGTGGACAGAGAGAGAGAGAGAGAGAGAGAGAG 351  
|||||  
DB 1598 AGCAGAAATTTCTCTCCAGATCCAAAGAAAGAAATATAAACCCTCAGATTTCAAG 1657  
QY 352 TATGAGCAAAAACACCAAAATGAGCTTACCTCAATCCAGAGATGCTTAAGTTGCA 411  
|||||  
DB 1658 AAGATAGCTAATATATATCTATCTGCTGCTTCTGATATCTGATATCTGATATCTGAT 1717  
QY 412 CTTGGAATACAGTCT-----GCAAGCAAGCTCAAAACCTTGAGCTGAGCTGTA 462  
|||||  
DB 1718 CTTGGAATACAGTCT-----GCAAGCAAGCTCAAAACCTTGAGCTGAGCTGTA 1777  
QY 463 CGTCCCGCTGCTCAATGCTACCTGCTCAATGCTCAATGCTCAATGCTCAATGCTCAATG 522  
|||||  
DB 1778 CGTCCCGCTGCTCAATGCTACCTGCTCAATGCTCAATGCTCAATGCTCAATGCTCAATG 1837  
QY 523 ACACAGCAAAATCCCAAGCCCTCTGAGTATTAACAAATATGCTCCCAAGAGAGCCCTCC 582  
|||||  
DB 1838 ATACAGCAAAATCCCAAGCCCTCTGAGTATTAACAAATATGCTCCCAAGAGAGCCCTCC 1897  
QY 583 CATTCCTAGGGCCAGTGCACAGCCCTGATTAATCAATGATTAATGATTAATGATTAATG 642  
|||||  
DB 1898 CATTCCTAGGGCCAGTGCACAGCCCTGATTAATCAATGATTAATGATTAATGATTAATG 1957  
QY 643 TGAATGAG 702  
|||||  
DB 1958 TGAATGAG 2017  
QY 703 GGTATTTCAAACTATGACAGAAATGTAATGATTAATGATTAATGATTAATGATTAATG 762  
|||||  
DB 2018 GGTATTTCAAACTATGACAGAAATGTAATGATTAATGATTAATGATTAATGATTAATG 2077  
QY 763 GGTATTTCAAACTATGACAGAAATGTAATGATTAATGATTAATGATTAATGATTAATG 822  
|||||

Db 2078 GAAACACACGCGTAGGCTAGTTTAAACACACAGACAGAACAAACCCCTGTATATACCG 2137  
 QY 823 GCTGGATGAGAAATGATGAAATACAAATGAAATCCACCAAGACCTGAATATATAAGATG 882  
 Db 2138 GCTATATTGAAAATGTAATTTACTGAAACCCACCAAGACCTGAATATATAAGATGACC 2197  
 QY 883 ATGTTCACACAGCAAGCTGTGTTTACAGACAGACATCTCGGAGGCTGATTTGTGCTT 942  
 Db 2198 TGGCAAAAGCTGAATAGAGACTTTACAGACTTAACCTCGAGAGGCTGATTTACTGTAT 2257  
 QY 943 CTGATGTCCCAATCTCCCACTGATCTCTTCCCACTGGCCAAATCAACCCGACCTGA 1002  
 Db 2258 CAGAGGCTCTCTCTCTGATATACACCTCTCTGCTCCACCAATTAATATATAGACCTTG 2317  
 QY 1003 AGCGGGAATTCAGCGGGGCACTCTATATCTGACTTGAACAGCTCTGGGATGAT 1062  
 Db 2318 AAGCAAAATTTAAAGAG---ATCACATTTCACTTTCAATGACACCCCTGGGATGCTC 2374  
 QY 1063 ATGACCATGGAACAGCTCAGATATATCATTTGAAATAGTAAAGTATCTTCTGATCTCA 1122  
 Db 2375 TAGATTAAGGAAGCAACAGCTACATATATAGATAGATTTCTGATCTCC 2434  
 QY 1123 GAGACAAATTCATATATCTCTCAAGTAAATGATCTGCTCATCCCAAGAGAGCA 1182  
 Db 2435 AAAAAAATTTGACATGCTACTTATGATGATCTTCACTTAACCTTAAGAGAGCCG 2494  
 QY 1183 ACTGTGAGGAGCTTTTGTGTTAAACAGAAATGATCTTTGAAATGAGCAGATC 1242  
 Db 2495 GCTCAGATGAAATTTTGAATTTAAACGAAACCTTTGATGATAGAAATGACCAACT 2554  
 QY 1243 TTTTCATGCTATTTAGGCTGTTGATAGCTGATCTGAAATCAAGAAATATCCACATTTG 1302  
 Db 2555 TGTATATTGCACTGCAAGCATCAATGAAGCCATCTCAGGAGGTTTCTAACATCG 2614  
 QY 1303 CACGAGTACTTGTATTCTT 1325  
 Db 2615 CACAAGCAATCAAGTTATCTT 2637  
 RESULT 7  
 AF001261 3317 bp mRNA MAM 19-NOV-1997  
 LOCUS AF001261 Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA, complete cds.  
 DEFINITION AF001261  
 ACCESSION AF001261  
 NID 92623762  
 VERSION AF001261.1 GI:2623762  
 KEYWORDS  
 SOURCE Bos taurus.  
 ORGANISM Bos taurus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
 REFERENCE 1 (bases 1 to 3317) Gruber, A.D., Abdel-Ghany, M., Levine, R., Elble, R.C., and Pauli, B.U.  
 AUTHORS Goodwin, A., and Pauli, B.U.  
 TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3317)  
 AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R., Goodwin, A., and Pauli, B.U.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA  
 FEATURES  
 SOURCE Location/Qualifiers  
 1..3317  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /cell\_type="aortic endothelial cells"  
 /clone="1"  
 /tissue\_type="Lung"  
 63..2780  
 CDS

/note="endothelial adhesion molecule; chloride channel"  
 /product="Lu-ECAM-1"  
 /protein\_id="AA86529.1"  
 /db\_xref="pid:92623763"  
 /db\_xref="gi:2623763"  
 /translation="MTICNLTFLTLHLPGKSSMVLINNGYGVIAINPSVE  
 DEKLINIKEMTENTSTYLFPAHTRRVYRNVSLIPMPKRSKEFFIKPSYQAD  
 VIVANPLYVGDPTLYQKGEKSKTHFIPNLINPHTIGSGRVYPHENAH  
 RMGIDPEYVNDQPEYISRNNTLEATRCSTHIEINVEKRCPSGCTSLCRDSOTG  
 LYEAKTFPKRSQAKESINMPSLSHYTECTERTHTETAENLONKCNKSTWD  
 INMSVDFONTSPKTEENPHTPELSLKSKORVCLVLDKSGMSHEDLFPQNAE  
 LYIOLYIEKSIATIEREDGHOVLELDNGAGLDYKNDGISRYETDYGNGRSLSK  
 VHAQARNNTARLNLRQPNKVLVYVPEYENGKILNPPREYVDLAKKIDFSRLT  
 SGGSFVSGAPRPENPVSYPSPKIDLDLAKREKEDYIOLSMTPAGVLDKGRANSYII  
 RISKSTMDREDNDNLVNTSLNLIKREKSGKENPEFKEHREVGCTFYISVQAIN  
 EANLISEVSHIVQAIKFIPLPDSVHDLDITKISEITLALGLPMIFSV"  
 BASE COUNT 1084 a 676 c 634 g 923 t  
 ORIGIN  
 Query Match 24.6%; Score 414.4; DB 3; Length 3317;  
 Best Local Similarity 59.6%; Pred. No. 2.9e-90;  
 Matches 804; Conservative 0; Mismatches 516; Indels 30; Gaps 5;  
 QY 1 AACAAAGTGGGCGCATATCCACACAGTGCCTTGGGGCCCTGACAGCTCAGAACTAG 60  
 Db 1351 AAGCAAGTGGGCGCATATCCACACATTCGCTCGGAGCCCTGCTGCCAAGAACTGG 1410  
 QY 61 AGAGCTGTCCAAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTGAGACA 120  
 Db 1411 AGACATTTGCAAAATATGACAGAGAGATATC-----GTTTTTTGCCAATTAAGACATAA 1464  
 QY 121 ATGGCTCATGATGCTTTTGGGGCCCTTCATCAGAAATGAGAGCTGTCTCAGGGCT 180  
 Db 1465 CTGGCTTACTAATGCTTTCAGTAAATTCATCTAAGATGAGACATCTCTACGACAG 1524  
 QY 181 CCATCAGCTTGAAGATGAGGATTAACCTCCAGAACAGCCAGCTGATGATGACAG 240  
 Db 1525 CTATTCAGTTGGAAGCAACACCTTGAATTAACAGAGGAAAGTAAAGCGCACAG 1584  
 QY 241 TGATGTGAGACAGCCGTGGGAAAGGACACTTTGTTCTTATCACTGCAACACGAC 300  
 Db 1585 TGCTGTAGACAGTACAGTTGGAATGACACTTCTTTGTTGTCACATGACCAATACAAA 1644  
 QY 301 CTCGCCAATCTCTCTCGGATCCAGTCCAGTGACAGAGCA-----AGGTGGCTTTG 351  
 Db 1645 AACCAGAAATGTTCTCCACATCCAAAGGAAAGAAATATAAACTCGATTTCAAG 1704  
 QY 352 TAGTGGCAAAAACACCAAAATGGCTTACCTCAATCCAGGAGCTGTAAGTTGGCA 411  
 Db 1705 AAGATAATTAATTTGATGCTGTGCTGCAAAATCTGATATGCAAGACAGGTA 1764  
 QY 412 CTGGAATTAACAGTCT-----GCAAGCAGCTCAAAACCTTGACCTGACTGCA 462  
 Db 1765 CTTGACTTACAGCTCTTAATATATATGACAGCTCTCAAAATGCTAATCAAGTACAGGA 1824  
 QY 463 CGTCCGCGGCGCAATCTACCTCCCTCCCAATTAAGTACTTCCAAAGCAAG 522  
 Db 1825 CCAGCTGAGCAAGAGTCTTACTATACCCAGTATGCAACAGCTCAGATGAGTAC 1884  
 QY 523 ACACCAAGCAATTTCCCAAGCCCTGTAGTTATGCAAAATATTCGCCAGAGACCTCC 582  
 Db 1885 ATACAGACATATCTTACGCCCAATGATGTTTATGCAAGTCAAGTCAAGGTTTTCG 1944  
 QY 583 CAATTCAGGGCCAGTGTACAGCCCTGATGATCAAGTGAATGGAAGAAAGTATCT 642  
 Db 1945 CTGTACTGGAGTACAGTGTATAGCCATTATGAAGAACGGAAGATGACATCAAGTACAT 2004





```

Db 1631 GAATCATCTTCCAGATCCAAAAGAAAAATATACACCTCAGATTCCAGATGAT 1690
Oy 357 GACAAAACACCAAAATGGCTACTCCAAATCCAGCATTTCTAAGTTGGCACTTG 416
Db 1691 AAACATAACACTCCGGTCTAGACTCAAAATACCGGCGACTGACAGACAGATGACTGG 1750
Oy 417 AATACACGCT---GCAAGCAAGCTCACAACCTTGACCTTACCTGACGTCGGTGG 473
Db 1751 ACTTACAGATCAGCGGTACCAAGTCTCGTTGATTACATGACATGACCATCTCAGCA 1810
Oy 474 TCCAAATGCTACCTGCTGCTTACAGATGATGATGATGATGATGATGATGATGATGAT 533
Db 1811 ACAAAGTCCCAACCATGAGACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1870
Oy 534 TTCCCAAGCCCTCTGTGATTTATGCAAAATATTCGCAAGAGCCCTCCCAATTTCTCAG 593
Db 1871 TACCTAGCCGATGATTTGTACGACGCGGTACGCAAGGATTTTGTGCTGCTGGGA 1930
Oy 594 GCGAGTGTACAGCCCTGATTTGATGATGATGATGATGATGATGATGATGATGATG 653
Db 1931 GCGAATGTACAGCCCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1990
Oy 654 GATATGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 713
Db 1991 GCAATGGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2050
Oy 714 ACTATGACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 773
Db 2051 GATTATCATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2110
Oy 774 GCGAAGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 833
Db 2111 ACCAAGCTGAGCTTAAGA---CAGAAGAACAAAGTCTTTATATATACCTGCTATG 2167
Oy 834 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 893
Db 2168 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2227
Oy 894 AAGCAAGTGTGTTGACGACGAACATCTCGGGAGGCTGCT---TGTGCTTCTGATGTC 950
Db 2228 ACAGTGAAGACTTCAACAGATTAACCTCTGGAGGCTGCTTACGCTGCTGAGAGGCC 2287
Oy 951 CCAATGCTCCCATACCTGATCTTCTCCACCTGGCCAAATACCGACCTGAGAGCGGAA 1010
Db 2288 CCGATGGGACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2347
Oy 1011 ATTCAGGGGGAGTCTCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1070
Db 2348 TTATAGG---ATTATATACCTTACATGATGAGGCGCCCTGGCAAGTTCTCGACAT 2404
Oy 1071 GGAACAGTCAACAGTATATCATGATGATGATGATGATGATGATGATGATGATGATG 1130
Db 2405 GGAAGAGCACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2464
Oy 1131 TTCAATGAATCTCTCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 1190
Db 2465 TTTAACAATGCTACTTGTAGTGAATGATGATGATGATGATGATGATGATGATGATG 2524
Oy 1191 GAAGCTTTTGTGTTAAACCAAAACATTAATTTGAAATGGGACANATTTTTCAT 1250
Db 2525 GAAACATTTAAATTTCAACACCAAAACATTTAAATGCAAAATGGGCTCCAGCTTACAT 2584
Oy 1251 GCTATTCAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1310
Db 2585 GCAATTCAGCGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2644
Oy 1311 TCTTTGTTTATCTCTCACA 1330
Db 2645 GTCAAGCTTACTTCTCTAGA 2664

```

RESULT 10  
AF043976

```

LOCUS AF043976 3415 bp mRNA PRI 07-APR-1999
DEFINITION Homo sapiens CLCA homolog (hCLCA3) mRNA, complete cds.
ACCESSION AF043976
VERSION 94572288
KEYWORDS AF043976.1 GI:45722288
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3415)
AUTHORS Gruber,A.D. and Pauli,B.U.
TITLE Molecular cloning and biochemical characterization of a truncated,
SECRETED member of the human family of Ca2+-activated Cl- channels
JOURNAL Biochim. Biophys. Acta 1444 (3), 418-423 (1999)
MEDLINE 99196715
REFERENCE 2 (bases 1 to 3415)
AUTHORS Gruber,A.D., Elble,R.C. and Pauli,B.U.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
FEATURES
SOURCE 1..3415
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3415
/gene="hCLCA3"
19..807
/gene="hCLCA3"
/codon_start=1
/product="CLCA homolog"
/protein_id="AAD3734.1"
/db_xref="PIR:G4572289"
/db_xref="GI:4572289"
/translation="MVESEKIVLILSLSPVLSVLTNNNGYDIVAINPSVE
DEKILNKEMVETASHTLPHATKORAFYRVSILIMTYKSEKVEYLIPKQETVDOAD
VIVADLVKRGDDPYTLOYGCGKRGQVHTPTPLNLNLTATGPKRKYVHRAHL
RMGVDFEINDDQPIYISRNTEFRSTRITVIVNVECGASCLARPERDSQTL
YEACFTFPRKSQTAKEISIVEMQMLDSVTECTEKTINKENPND."
BASE COUNT 1149 a 625 c 630 g 1011 t
ORIGIN
Query Match 21.8%; Score 366.6; DB 42; Length 3415;
Best Local Similarity 59.1%; Pred. No. 1e-78;
Matches 734; Conservative 0; Mismatches 479; Indels 30; Gaps 5;
Oy 115 AGAACAATGGCTCATGATGCTTTGGGCGCTTTCATCAGGAATGAGCTGCTCTC 174
Db 1469 ACATAAATGGCTTTATGATGCTTTCACAGAAATTCATGAAAGTGCGACATCTCTC 1528
Oy 175 AGCGCTCCATCCAGCTTGAGATGAGGATTAACCTCCAGAACGCCAGTGATGAATG 234
Db 1529 AGCAGCTCTTCAAGTTGGAATTAACCTTGAATATCCAGCGAAGAAATGATTAATG 1588
Oy 235 GCACAGTATCGTGACAGCACCGTGGAAGACACTTGTCTTATCAGCTGAGCAA 294
Db 1589 GTACAGTGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1648
Oy 295 GCGAGCTCCCAATCTCTCTGCGATCCAGTGGACAGAGA-----AGGTG 345
Db 1649 TACAAGACCCAGCAATATTTCTCAAGATCCAAAGAAAAAAATATCTACCTCAGATT 1708
Oy 346 GCTTGTAGTGGACAAACCAAAATGCGCTTACCTCAATCCAGGATGCTAAG 405
Db 1709 TTCAAGAAAGTGAATTAATATGCGTCTGCCGCTTCGATATACAGATATGACAGA 1768
Oy 406 TTGCACTTGGAAATACAGTCTGAG-----CAAGCTCACAAACCTTGACCTGA 456
Db 1769 CAGGCACTTGAGCTTACAGCGTTGCAACAAATCATACCAAAATCTCAATGCTTAC 1828
Oy 457 CTGTACGTCGCTGCGTCCATGCTACCTGCTCCTCATTTACAGTACTTCCAAAAGA 516

```

Db 1829 CAATGACCACTGAGCAGAGAGCCCTACACACTCCAGTATTCAGACTGCTCATGA 1888  
 QY 517 ACAGAGACACAGCAAAATCCCCAGCCCTCTGGTAGTTTATGCAAAATTTGCCAAGAG 576  
 Db 1889 GTCAAAATACAGCTCATATACCTACCCAGATGTTTATGATGATGATGATGATGAT 1948  
 QY 577 CCTCCCAATTCCTCAGGGCCAGTGTACAGCCCTGATGATGATGATGATGATGATGATGAT 636  
 Db 1949 TTTCTTCTGTTCTGGGAATTCATGTAACGCCATTATGAAATAGAGAGGACTTCAG 2008  
 QY 637 TTACCTTGAATCTACTGATATATGAGCAGAGTGTCTGATGCTACTAGAGTACGCTGTCT 696  
 Db 2009 TAACATTGAGCTCTGACACATGCGCAGAGTGTCTGATGCTGATGATGATGATGATGAT 2068  
 QY 697 ACTCAGATATTTACACAACTTATGACACGAATGTAGATACAGTGTAAAGTGGGGCTC 756  
 Db 2069 ACTCAGATATTTATACAACTTATGACATGATGATGATGATGATGATGATGATGATGAT 2128  
 QY 757 TGGGAGGATTAAGCAGCAGAGAGAGATGATACCCAGAGAGTGGAGCAGCTGTACA 816  
 Db 2129 AGGCAAGAAAAACACAGCTAGGC-----TAACTCAACAAAGATTAAGCTGTATG 2182  
 QY 817 TACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876  
 Db 2183 TACCGCGCTATGCTGAATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2242  
 QY 877 AGATGATGTTCAACACAGCAAGTGTCTTTCAGCAGAACTCTCGGAGGCTCATTTG 936  
 Db 2243 ATGATGTTGGAAGAGGCTCAAAACAGCACTTCAAGAGCTCACTCTGAGGCTGTGTTA 2302  
 QY 937 TGGCTTCTGATGT--CCCAATGCTCCCATACCTGATCTCTCCAGCTGGCCAAATCA 993  
 Db 2303 CTGATCTAGAGAGTGCCTCTTATGATGATGATGATGATGATGATGATGATGATGATG 2362  
 QY 994 CCGAGCTGAAGCGGGAATTCACGGGGGAGTCTCATTAATCTGACTTGAACAGCTCTG 1053  
 Db 2363 TAGACCTGAGGCTTAAGTTTCAAGAG--ATCATATCAACTTTCATGAGCTGCCCTG 2419  
 QY 1054 GGGATGATATGACCATGGAAGAGCTCACAAGTATATCATTCGATGATGATGATGATGAT 1113  
 Db 2420 GCAAGGCTCTGATTAAGAAAGAGTGAAGCTGATGATGATGATGATGATGATGATGATG 2479  
 QY 1114 TTGATCTCAGACAGCAAGTTCATGATGATGATGATGATGATGATGATGATGATGATG 1173  
 Db 2480 TGGACCTTCAAGAGATTTTATTAAGCTGTTTATTAATTAATTAATTAATTAATTAATTA 2539  
 QY 1174 AGAAGCCCACTGAGAGAGTCTTTTGTGTTTAAACAGAAACATTTCTTTGAAATG 1233  
 Db 2540 AGGAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2599  
 QY 1234 GCACAGATCTTTTCTATGCTTATTCAGGCTGTGATGATGATGATGATGATGATGATGAT 1293  
 Db 2600 GTACGACATTTTATTAATGCAATTCAGGCAATTCAGGCAATTCAGGCAATTCAGGCAAT 2659  
 QY 1294 CCACATGTCAGAGATCTTTGTTTATTTCTCCACAGACTCC 1336  
 Db 2660 CAACATGTCAGAGCACTTAATTTCTCTCCAGAGAAC 2702

RESULT 11  
 AF001263 2820 bp mRNA 19-NOV-1997  
 LOCUS Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF001263  
 NID 92623766  
 VERSION AF001263.1 GI:2623766  
 KEYWORDS  
 SOURCE Bos taurus.  
 ORGANISM Bos taurus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 2820)  
 AUTHORS Elble,R.C., Widom,J., Gruber,A.D., Abdel-chany,M., Levine,R.,  
 TITLE Goodwin,A. and Pauli,B.U.  
 JOURNAL Cloning and characterization of Lu-ECAM-1 suggest it is an  
 REFERENCE endothelial chloride channel  
 AUTHORS Unpublished  
 2 (bases 1 to 2820)  
 Elble,R.C., Widom,J., Gruber,A.D., Abdel-chany,M., Levine,R.,  
 TITLE Goodwin,A. and Pauli,B.U.  
 JOURNAL Direct Submission  
 Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
 Ithaca, NY 14853, USA  
 FEATURES  
 source location/Qualifiers  
 1..2820  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone="3"  
 /tissue-type="lung"  
 /cell-type="aortic endothelial cells"  
 194..2856  
 /note="endothelial adhesion molecule; chloride channel"  
 /codon\_start=1  
 /product="Lu-ECAM-1"  
 /protein\_id="AAB86531.1"  
 /db\_xref="pid:62623767"  
 /db\_xref="gi:2623767"  
 /translation="NYCLNVLILFLHLPLPKMSVMNLINNGYGVIAINPSVE  
 DEKLINIKEMVEASTLYLFHATKRVYRNVSLIPMTWKSXEYFIRKQESYDAD  
 VIVANVYLKRGDPPYLYOGRGCKYHIFENFLPNTNFHYSGRGVFVEMAH  
 RMGIDFEDVNDOPFYSRKNTIEATSCSHITGIVNFKKPGSCITSCRDSDGT  
 LYAKCTFLPKSOTAKESIMEPKSLHSEYETETKTNTEANLONKMGKSTDV  
 IINSDPONTSPMTENPHTPEFSLKRLORVYCVLDKSGMSMEDLPONMAE  
 LTIQYIEKSLVGVATFDSVAETIOLNHLRTIDDNVYQKTKLPPVANGSTICGL  
 KAFQAIINSDOSTGSEIILLTGDNDENISCFEDVKRSGLIHTIILAGPSAKLE  
 TLSNMTGGRFEFANKDITGLTNAFSRISNSGSIITQAIQLESKALKITGRKRVNTV  
 PVDSTVGNDFEVVWTIOKPEIYLDPPKRGKISDFEKDINISARLQIPGNET  
 GWTYVSLNNHASSOMLVTYVTRASPPIPIVIAHNSOHTAHPSPMIVYAVSO  
 GELPVGISVIAIETREDGHOVTEIEMDAGDPTKNDGISYRYTIDVYGNRSYELK  
 YHAQNNNTARLRLROPKNKLVLPVPEYVENGKITLLPPEYVADDLAKKIDFSLT  
 SGGSTFVSAGAPPENHVSFPPSKITIDLAKEFEDYIQLSWTAPGNVLDKGRASGSP  
 MSRFSGVAKVELIQLQHSFO"  
 BASE COUNT 940 a 565 c 560 g 755 t  
 ORIGIN  
 Query Match 19.0%; Score 320.2; DB 3; Length 2820;  
 Best Local Similarity 59.2%; Pred. No. 1.7e-67;  
 Matches 651; Conservative 0; Mismatches 418; Indels 30; Gaps 5;  
 QY 1 AACAAAGTGGTGCATGATCCACACAGTGGCTTGGGGCCCTCTGCAGCTCAGAACTAG 60  
 Db 1482 AACGAAGTGGTGCATGATCCACACAGTGGCTTGGGGCCCTCTGCAGCTCAGAACTAG 1541  
 QY 61 AGGAGCTGTCCAAATGACAGAGGTTTACAGACATGCTTACAGATCACTTACAGACA 120  
 Db 1542 AGAATGTGTAATTAATGACAGAGGATATC-----GTTTTTTGCCAATAAGACATA 1595  
 QY 121 ATGGCTCATTGATGCTTTTGGGGCCCTTTCATCAGAGAAATGAGCTGTCTCAGCGCT 180  
 Db 1596 CTGGCCTTACTAATGCTTTCAGTAGAATTTCACTGAGTAAGAGCATCTCAGAGG 1655  
 QY 181 CCATCCAGCTTGAGAGTAAGGATTAACCTCCAGAAACAGCAGTGTGATTAATGGCAGAG 240  
 Db 1656 CTATTCAGTTGGAAGAGAAAGCCCTTGAAATTAACAGAAAGAAAGATTAAGCGCAG 1715  
 QY 241 TGAATCGGACAGACACGCTGGGAAAGCACTTTTCTTATCACTGAGCAAGCAGC 300  
 Db 1716 TGCGCTGAGACAGAGAGTGGAAATGACACTTCTTTGTGTCACATGACATACAA 1775  
 QY 301 CTCGCCAATCTCTCTGCGATCCAGTGGACAGACG-----AGTGGCTTTG 351  
 Db 1776 AACCGAAATGTTCTCCAGATCCAAAGGAAAGAAATTAACCTCGATTTCAAG 1835





QY 463 CTTCCCGTCCGATGCTACCTGCTCCATTAAGTGTGCTTCCAAAGCAAGG 522  
Db 1821 CCACGCGAGCAGAGAGTCTACTATACCCCGATATGCAACACTCAGAGTCAAC 1880  
QY 523 ACACGAGCAATATCCCGAGCTGTAGTATGCAAAATTTGGCCAGGAGCTCC 582  
Db 1881 ATACGACGATTAATCTTACGCCAATGATGTTTATGCAAAATCAGTCAAGGTTTTC 1940  
QY 583 CAATTCAGGCGGAGTGTACAGCCCTGATTGAATGATGTAAGTGAAGAAACAGTTACT 642  
Db 1941 CTGATCTGGAATCAGTGAATATGACCATTAATGAAACGAGATGACATCAATACAT 2000  
QY 643 TGGAACTACTGATATATGAGAGAGTGTGATGCTACTAAGATACAGTGTCTCTCAA 702  
Db 2001 TGGAGCTCTGGACAAATGATGAGGTCGATGATGATGATGATGATGATGATGATGAT 2060  
QY 703 GGTATTTCAACAATCTATGACAGATGATGATGATGATGATGATGATGATGATGATGAT 762  
Db 2061 GATATCTTACAGATTAATGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2120  
QY 763 GAGTTAAGCAGCAGCAGAGAGATGATACCCAGCAGAGTGGAGCAGTATACATCTG 822  
Db 2121 GAACACACAGCGCTAGGCTAAATTAAGACACACAGACAAAGTCTATATGTTCCAG 2180  
QY 823 GCTGATTTGAGAAATGATGAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 882  
Db 2181 GCTACGTTGAAAGAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2240  
QY 883 ATGTTCAACACAGCAAGTGTGTTTACAGCAAGCAATCTCGGAGGCTCATTTGGGCTT 942  
Db 2241 TGGCAAAAGCTAAATTAAGAGCTTTAGCAGAGCTAACCTCTGAGAGGCTATTTAGTAT 2300  
QY 943 CTGATGTCTCCAAATCTGCCA---TACCTGATCTCTCCCACTGCGCAATCAGCAGC 999  
Db 2301 CAGGAGCT 2360  
QY 1000 TGAAGGCGGAATTCAGGCGGAGCTCTCATTAATCTGATGAGTGTGAGGAGTGTGAGG 1059  
Db 2361 TTGAGGCTTAAGTTCAGAGAG---ATTATATTAATTAATTAATTAATTAATTAATTAAT 2417  
QY 1060 ATTATGACCATGAGACAG 1077  
Db 2418 TCCATAGATTAAGAGAAAG 2435

RESULT 13  
AB026833 3604 bp mRNA PRI 26-MAY-1999  
LOCUS Homo sapiens mRNA for chloride channel protein, complete cds.  
ACCESSION AB026833  
NID 94887600  
VERSION AB026833.1 GI:4887600  
KEYWORDS chloride channel protein.  
SOURCE Homo sapiens adult corneal epithelium cDNA to mRNA,  
clone, lib:lambda ZAP clone:lambda 13802.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Itoh, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.  
TITLE Isolation and characterization of chloride channel in human corneal  
epithelium  
JOURNAL Unpublished (1999)  
REFERENCE 2 (bases 1 to 3604)  
AUTHORS Itoh, R., Kawamoto, S. and Okubo, K.  
TITLE Direct Submission  
JOURNAL Submitted (28-APR-1999) to the DDBJ/EMBL/Genbank databases. Reiko  
Itoh, Institute for Molecular and Cellular Biology, Osaka  
University, Yamada-oka 1-3, Suita, Osaka 565-0871, Japan  
(E-mail:reiko@imcb.osaka-u.ac.jp, Tel:81-6-6879-7992,  
Fax:81-6-6877-1922)  
FEATURES  
Location/Qualifiers

source 1..3604  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="lambda 13802"  
/clone\_lib="lambda ZAP"  
/dev\_stage="adult"  
/map="1p32"  
/tissue\_type="corneal epithelium"  
42..2873  
/codon\_start=1  
/product="chloride channel protein"  
/protein\_id="AA07810.1"  
/db\_xref="pid:d1041578"  
/db\_xref="pid:g4887601"  
/db\_xref="gi:4887601"  
/translation="MTORSIAGPICNLKFEVTLVLSSELPFGAGVQLODNGNGL  
IAINPOYEMONLISNIKEMITEASFLEFATRRVFERIKILIPATWANNRSKR  
OESYERANVIVTMWYGAHGDDPYTLIOGRCGKRGKYLHPNPLINDNTLAGSGR  
VFHERAHLMGCFDEYINDKPYINGONDIKVTROSSDITGIFVECKGPCDNCIT  
SKLKEGCTFTYSTONATASIMFWSLSSVFEFCNASTHNOEAPMLQNMCSLRSAW  
DYITDSADFHSHSPNGTELPPTFSLVQAGDVVCLVDVSKMAEADRLQLODA  
AEFYLMQIVAEIHFTVGIASPDSKGEIRAOHQINSDDKRLVSYLPTVSAIDISI  
CSGLKGEFEVVKLNGAGSVMLIVTSQDDKILGNCLPVLSSGTHSIAIGSSA  
PNLELSRLTGLGKFPYDPSISNSNMDAFSSGTDIFQOHIQLESTGEVXPH  
QLENTYVDNVTGNDTMEIYTMQASPEIILDPDGRKYTNNTFTNFTIRASIMI  
PGTKAPRHWITLNTNTHSLQALKLVTSRASSAVPAPVTEAFVERDSLHPFPMVI  
YANVKGFPYILNAVATVAIEPPTGDPVTLRLDDAGAVANDISYKTFPSFAN  
GRYSLKRVHNSPISIPASHIPGSAMVPGYTANGONMAPRSVGNEDERWAG  
FSYVSGSGSFSVGVPAHPDVPCKIILDEAVVEEELTSLMTPAGDPQOGAT  
SYEIRMSKSLQNDOPFNALIVNYSKRNPQAGIREIFTSPOISNPGEPHONET  
HSHRIYVAIRAMDNRNLSQASVAINQAPLPIPNSDPVPARYLLIKVLTAMGLIG  
IICLIYVTHHTLSRRKADKRENGIKLL"  
BASE COUNT 1113 a 740 c 732 g 1017 t 2 others  
ORIGIN

Query Match 17.9% Score 301.6; DB 9; Length 3604;  
Best Local Similarity 55.6%; Pred. No. 5.4e-63;  
Matches 693; Conservative 0; Mismatches 529; Indels 24; Gaps 5;

QY 6 AGTGTGCGCATCATCACACAGTGCCTTTGGGCGCTCTGCGAGCTCAAGACAGAG 65  
Db 1347 AGTGTGCGCATCATCACAGTGCCTTTGGGCGCTCTGCGAGCTCAAGACAGAG 1406  
QY 66 CTGTCCAAATGACAGAGAGTTCACACATATGCTTCACATCAAGTTCAGACATGCG 125  
Db 1407 TTATCAGGCTTACAGAGAGGTTTAAAGTCTTCTTCCAGATATATCAAACTCAATAGC 1466  
QY 126 CTCATGATGCTTTTGGGCGCTTTCATCAGGAATGAGAGTGTCTCTCAGCGCTCATC 185  
Db 1467 ATGATGATGCTTTTGGGCGCTTTCATCAGGAATGAGAGTGTCTCTCAGCGCTCATC 1526  
QY 186 CAGCTTGAGTAGGAGGATTAAACCTCCAGAACGCCAGTGAATGAGAGTATC 245  
Db 1527 CAGCTTGAGTAGGAGGATTAAACCTCCAGAACGCCAGTGAATGAGAGTATC 1586  
QY 246 GTGACAGACCGGTGGGAAAGAGACGTTGTTTATACCTG---GACACGAGCCT 302  
Db 1587 GTGACAGACCGGTGGGAAAGAGACGTTGTTTATACCTG---GACACGAGCCT 1646  
QY 303 CCCCAATCCCTCTCTGCGGATCCAGTGGACAGAGAA-----AGTGGCTTTGTAGTG 356  
Db 1647 CCTGAGATTAATTAATTTGATCCGATGAGAGGAAATATACACAAATTAATTTATACC 1706  
QY 357 GACAAACCAACCAATATGCTTACCTCAATCCAGGACATGCTAAGTGTGACCTTGG 416  
Db 1707 AATTAATCTTTGGAGACGATGCTTTGATTCAGAGACAGTAAAGCTGGGACATGG 1766  
QY 417 AAATACAGTGTGACAGAGCTCA-----CAAACTTGACCTGACTGTACGCTC 467  
Db 1767 ACTTACACCGCTGAACATATCCATCATCTCTGCAAGCCCTGAAGAGTACAGATGACCTCT 1826







```
HKASNDKSVKMAVIMTIFSFMAESPGIITHYOFYDYSSIIICVSKSLDKLFIJLS
ILNATVHCYISFRVSTQYREVAKKLFSPFIIIRRTETALAIKNSIHAVETPTSGKSSST
KTVVSRGPRLEKIHIFHTATMTIMKLERDVRLLLFERLGHSAEAEARNICG
AGEEGALSYNTAKSWFOKFNDFSLSEIERSGRPVELNEEDLVKLYEERPLSLRN
EKKLECHSTIARHLRGLFTSKLIGWVHELASQKLRVAVCTDILTEFRKFDLM
NLVTGDEKVVLYVNSHRKRWLPIGEGKIPTRPDLPKIKICVMWGVGVPMVELL
PINKTITADYCAOLDRAVEKTKNGKYEKLPFLHDNARPHVAKKTPOKLODGLGWTLP
PPYSPDLATDVLHPLSLSDYMRDKOFDDEHKTLELSTFFSSRSPDFFSRGIIMLPS
KMQQVVDINGETICE
complement(40651..40722)
/gene="Y37A1B.t1"
/Note="TGC Ala A-tRNA: predicted using tRNAscan-SE-1.11;
preliminary prediction: similar to tRNA-Ala"
complement(40651..40722)
/gene="Y37A1B.t1"
complement(42945..46379)
/gene="Y37A1B.9"
complement(42945..43220,43449..43516,43693..44020,
44065..44361,45584..45784,45832..46071,46346..46379))
/gene="Y37A1B.9"
/codon_start=1
/protein_id="CAA19492.1"
/db_xref="PID:e1350275"
/db_xref="GI:3880778"
/translation="MCKNIIDHYFPECTFNHSEVTSKTIKDFPSNCEIVCGILNFN
SNTDISEDLKRFQNLKELDDLVKENSNTLSLSEFSYDFNFCYTGISLIINSL
EYVDILKRAVYIRANEDIENFQIMNNTKLDASKLGNRMNDLELKVLGNLKDQC
FYDLANPNFDFETCKLYGGLNLNTIENSVNFSKFLGHNIEVAHTNFKDL
SEKLNLTOMENEVENEIIINHDPNEMTRLGWTSLEYKGRRTYIMNLENLHPD
FOLSLLELIEFSKIGVEFNIGATCODENFLDANYGRCOFOTLSNLGYCEIIGD
LYIDSEKYLEKLIHLPFEFTKIQYKSEATOPAITIRSNKTIKRARIRPLEHYI
SKTSDVAVNENPSMISNQECILRTIYQINLINRNLDCGWEPTLLEMLVFPDTN
VFNESDNKHNGLIFEMWLGINIKMFVF"
47044..51474
/gene="Y37A1B.8"
join(47044..47137,47693..47977,48348..48539,48590..48873,
```

Search completed: August 5, 1999, 23:24:22  
Job time: 8345 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:32 ; Search time 3059.95 Seconds  
(without alignments)  
141.818 Million cell updates/sec

Title: US-09-049-696-5  
Perfect score: 220  
Sequence: 1 CTAAGCTGAAATCTGTACA.....GACAAATCTGAAAGCATGGC 220

Scoring table: OLIGO\_NUC

Searched: 2546578 segs, 98626752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	220	100.0	480	35	AA581198
2	19	8.6	205	25	W11372
3	19	8.6	199	50	AV026729
4	18	8.2	427	34	AA498535
5	18	8.2	525	49	A1659206
6	17	7.7	443	20	W75917
7	17	7.7	520	22	R61896
8	17	7.7	428	22	R62089
9	17	7.7	412	25	N55346
10	17	7.7	220	25	N77352
11	17	7.7	260	26	W71854
12	17	7.7	292	27	C06739
13	17	7.7	213	29	AA142568
14	17	7.7	313	29	AA164101
15	17	7.7	339	29	AA166384
16	17	7.7	361	29	AA178674
17	17	7.7	516	29	AA190916
18	17	7.7	533	33	AA394843
19	17	7.7	429	33	AA414677
20	17	7.7	417	34	AA465108
21	17	7.7	434	34	AA488989
22	17	7.7	476	36	AA636732
23	17	7.7	320	36	C74266
24	17	7.7	525	37	AA696398
25	17	7.7	346	37	AA704201
26	17	7.7	372	38	AA744665
27	17	7.7	463	38	AA809480
28	17	7.7	379	39	AA813879
29	17	7.7	472	39	AA815232
30	17	7.7	363	40	AA971033
31	17	7.7	432	40	AA976402
32	17	7.7	485	41	A1056234
33	17	7.7	300	43	C98972
34	17	7.7	548	44	AU043085
35	17	7.7	386	46	A1451099
36	17	7.7	313	48	A1592108
37	17	7.7	388	48	A1592326
38	17	7.7	454	48	A1604060
39	17	7.7	412	50	A1674736
40	17	7.7	652	51	A1725132
41	17	7.7	481	51	A1733457
42	16	7.3	254	21	R13614
43	16	7.3	376	21	R17975
44	16	7.3	509	21	R19757
45	16	7.3	610	51	A1714744

ALIGNMENTS

RESULT 1  
LOCUS AA581198  
DEFINITION n38c07.r1 NCI-CGAP\_Col Homo sapiens cDNA clone IMAGE:802572 5' similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.  
ACCESSION AA581198  
NID g2358970

VERSION AA581198.1 GI:2358970  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 480)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 9, 1995 this sequence version replaced gi:803001.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies Inc., David Krizman,  
Ph.D.  
CDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 405.  
FEATURES  
source  
1. 480  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:802572"  
/clone\_lib="NCI-CGAP\_Col1"  
/tissue\_type="bulk tumor"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pCMV-SPORT2; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt."  
BASE COUNT 168 a 101 c 97 g 114 t  
ORIGIN  
Query Match 100.0%; Score 220; DB 35; Length 480;  
Best Local Similarity 100.0%; Pred. No. 8.7e-109; Indels 0; Gaps 0;  
Matches 220; Conservative 0; Mismatches 0;  
QY 1 CTATGATGTAATCTGTACAGAACAAACCAACAAAGAGCTCCAAACAGCAAAATC 60  
DB 215 CTATGATGTAATCTGTACAGAACAAACCAACAAAGAGCTCCAAACAGCAAAATC 274  
QY 61 AAAAATGCAATCTCCGAAGACATGGAGAGTGCCTGATTCGTGAGAGCTTTAAGAAA 120  
DB 275 AAAAATGCAATCTCCGAAGACATGGAGAGTGCCTGATTCGTGAGAGCTTTAAGAAA 334  
QY 121 CCACCTCTAAGAACACACACACCAATCCACCTCTCTATGCTGCACATGAGCAAA 180  
DB 335 CCACCTCTAAGAACACACACACCAATCCACCTCTCTATGCTGCACATGAGCAAA 394  
QY 181 GAATGTGTGTAGTCTTGACAAATCTGGAAGATGCG 220  
DB 395 GAATGTGTGTAGTCTTGACAAATCTGGAAGATGCG 434  
RESULT 2  
LOCUS W11372 205 bp mRNA EST 02-OCT-1997  
DEFINITION ma86602.r1 Soares mouse p3NMf19.5 Mus musculus CDNA clone  
IMAGE:317594 5' similar to SW:HB1\_MOUSE P02088 HEMOGLOBIN BETA-1  
CHAIN 1; mRNA sequence.  
ACCESSION W11372  
NID 91285738  
VERSION W11372.1 GI:1285738  
KEYWORDS EST.

SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 205)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellendberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:837405.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mousese@wustl.edu](mailto:mousese@wustl.edu)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
MG1: 208210  
Trace considered overall poor quality  
Possible reversed clone; similarity on wrong strand  
Seq primer: mob.REGA+ET  
High quality sequence stop: 1.  
FEATURES  
source  
1. 205  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="17q21"  
/clone="IMAGE:317594"  
/clone\_lib="Soares mouse p3NMf19.5"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT7b (pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand CDNA  
was primed with a Not I - oligo(dt) primer 15,  
TGTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3',  
double-stranded CDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."  
BASE COUNT 52 a 39 c 52 g 62 t  
ORIGIN  
Query Match 8.6%; Score 19; DB 25; Length 205;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 98 TGATTCTGAGACTTTAAG 116  
DB 169 TGATTCTGAGACTTTAAG 187  
RESULT 3  
LOCUS AV026729/c 199 bp mRNA EST 02-JUN-1999  
DEFINITION AV026729 Mus musculus adult C57BL/6J liver Mus musculus CDNA clone  
1300002D03, mRNA sequence.  
ACCESSION AV026729  
NID 94806339  
VERSION AV026729.1 GI:4806339  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert\_Strausberg@nh.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R  
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/dbfp/image/image.html

Seq primer: -40UP from Gibco  
 High quality sequence stop: 459.

## FEATURES

Source

Location/Qualifiers  
 1. 525  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2249902"  
 /clone.lib="NCI\_CGAP\_Pr28"  
 /sex="male"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)  
 with a modified polylinker; plasmid DNA from the  
 normalized library NCI\_CGAP\_Pr22 was prepared, and ss  
 circles were made in vitro. Following HAP purification,  
 this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (clones  
 985608-986759, 110192-110195, and 1217928-1220615)."  
 Subtraction by Bento Soares and M. Fatima Bonaldo."

## BASE COUNT

187 a 67 c 85 g 186 t

## ORIGIN

Query Match  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 47 AACACAGCAAAATCAAAA 64  
 Db 325 AACACAGCAAAATCAAAA 342

RESULT 6  
 M75917 443 bp mRNA EST 16-SEP-1992  
 LOCUS CEL06D10 Chris Martin sorted cDNA library Caenorhabditis elegans  
 DEFINITION cDNA clone cm06d10 5', mRNA sequence.

ACCESSION  
 M75917  
 NID 9275240

VERSION  
 M75917.1 GI:275240

KEYWORDS  
 EST.

SOURCE  
 Caenorhabditis elegans.

ORGANISM  
 Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditidae; Peloderinae; Peloderinae; Caenorhabditis.

REFERENCE  
 1 (bases 1 to 443)  
 Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A.,  
 Hillier, L., Durbin, R.K., Green, P., Showkneen, R., Halloran, N.,  
 Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J.,  
 and Sulston, J.

A survey of expressed genes in Caenorhabditis elegans

Nature Genet. 1, 114-123 (1992)

JOURNAL  
 MEDLINE  
 93250982

## COMMENT

Contact: Waterston R.H.(USA) and Sulston J.E.(UK)  
 (USA) Dept. of Genetics or (UK)  
 (USA) Washington Univ. School of Medicine or (UK) MRC Laboratory of  
 Molecular Biology

Box 8233,4566 Scott Ave., St. Louis, MI 63110, USA, or Hills  
 Road, Cambridge CB2 2QH, UK

Tel: (USA) (314)3627072 or (UK) (0223)248011

Fax: (USA) (314)3624137 or (UK) (0223)402008

Email: tw@genome.wustl.edu or jes@mrc-lmba.cambridge.ac.uk

single read.  
 Location/Qualifiers

Source

1. 443  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /clone="cm06d10"  
 /clone.lib="Chris Martin sorted cDNA library"  
 /lab\_host="MC1061"  
 /note="Vector: lambda phage SHX2; Mixed stage  
 hermaprodite cDNA library. Partially normalized by  
 successively picking groups of clones that didn't  
 hybridize to previously picked clones. Vector: lambda phage  
 SHX2 (Lipschitz, D.H. et al., Gene 88:25-36 (1990)) Host:  
 MC1061"

## BASE COUNT

126 a 107 c 91 g 117 t 2 others

## ORIGIN

Query Match  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 TTGAATCTGTACAGAA 23  
 Db 57 TTGAATCTGTACAGAA 41

RESULT 7  
 R61896 520 bp mRNA EST 31-MAY-1995  
 LOCUS EST004 coelomocyte Strongylocentrotus purpuratus cDNA 5' end  
 DEFINITION similar to arylsulfatase, mRNA sequence.

ACCESSION  
 R61896  
 NID 9832591

VERSION  
 R61896.1 GI:832591

KEYWORDS  
 EST.

SOURCE  
 Strongylocentrotus purpuratus.

ORGANISM  
 Strongylocentrotus purpuratus.  
 Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinodae;  
 Euechinozoa; Echinacea; Echinodae; Strongylocentrotidae;

REFERENCE  
 1 (bases 1 to 520)  
 Smith, L.C., Chang, L., Britten, R.J., and Davidson, E.H.

Sea urchin genes expressed in activated coelomocytes are identified  
 by expressed sequence tags. Complement homologues and other  
 putative immune response genes suggest immune system homology  
 within the deuterostomes

J. Immunol. 156, 593-602 (1996)

JOURNAL  
 MEDLINE  
 96132985

On May 9, 1995 this sequence version replaced gi:802365.  
 Other ESTs: EST003, EST072, EST401

Contact: Smith, L. Courtney  
 Laboratory of E.H. Davidson  
 California Institute of Technology

Division of Biology, Caltech, Pasadena, CA 91125

Tel: 8183954944

Fax: 8185838351

Email: court@caltech.edu

Insert Length: 1266 Std Error: 0.00

Seq primer: Sp6

High quality sequence stop: 308.

Location/Qualifiers

1. 520  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone.lib="coelomocyte"  
 /lab\_host="NM522"

/note="Vector: lambda ExCell; Site\_1: EcoRI; Site\_2: NotI;  
 mRNA was isolated from LPS activated coelomocytes. cDNA  
 was constructed with random primer containing a NotI site,  
 and an EcoRI adapter was ligated at the 5' end. The cDNA  
 was directionally cloned."

BASE COUNT  
 110 a 119 c 141 g 133 t 17 others

ORIGIN



Query Match 7.7%; Score 17; DB 22; Length 520;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 79 GCACATGGAGATGATC 95  
 |||||||  
 Db 254 GCACATGGAGATGATC 270

RESULT 8  
 R62089 428 bp mRNA EST 31-MAY-1995  
 LOCUS R62089  
 DEFINITION EST147 coelomocyte Strongylocentrotus purpuratus cDNA 5' end, mRNA sequence.  
 ACCESSION R62089  
 NID 9832784  
 VERSION R62089.1 GI:832784  
 KEYWORDS EST.  
 SOURCE Strongylocentrotus purpuratus.  
 ORGANISM Strongylocentrotus purpuratus.  
 Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea; Echinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.  
 1 (bases 1 to 428)  
 Smith, L.C., Chang, L., Britten, R.J. and Davidson, E.H.  
 Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. Complement homologues and other putative immune response genes suggest immune system homology within the deuterostomes  
 J. Immunol. 156, 593-602 (1996)  
 JOURNAL 96132985  
 MEDLINE  
 COMMENT On May 9, 1995 this sequence version replaced gi:8025399.

REFERENCE 1  
 Smith, L.C., Chang, L., Britten, R.J. and Davidson, E.H.  
 Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. Complement homologues and other putative immune response genes suggest immune system homology within the deuterostomes  
 J. Immunol. 156, 593-602 (1996)  
 JOURNAL 96132985  
 MEDLINE  
 COMMENT On May 9, 1995 this sequence version replaced gi:8025399.

CONTACT: Smith, L. Courtney  
 Laboratory of E.H. Davidson  
 California Institute of Technology  
 Division of Biology, Caltech, Pasadena, CA 91125  
 Tel: 8183954944  
 Fax: 8185838351  
 Email: court@mlrsky.caltech.edu  
 Insert Length: 1352 Std Error: 0.00  
 Seq primer: Sp6  
 High quality sequence stop: 136.  
 Location/Qualifiers  
 1..428  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone\_lib="coelomocyte"  
 /lab\_host="NM522"  
 /note="Vector: lambda Excell; Site\_1: EcoRI; Site\_2: NotI; mRNA was isolated from lps activated coelomocytes. cDNA was constructed with random primer containing a NotI site, and an EcoRI adapter was ligated at the 5' end. The cDNA was directionally cloned."

BASE COUNT 90 a 91 c 119 g 121 t 7 others  
 ORIGIN

Query Match 7.7%; Score 17; DB 22; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 79 GCACATGGAGATGATC 95  
 |||||||  
 Db 288 GCACATGGAGATGATC 304

RESULT 9  
 N55346 412 bp mRNA EST 01-FEB-1997  
 LOCUS N55346  
 DEFINITION yv47d02.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone

IMAGE:245859 3', mRNA sequence.  
 ACCESSION N55346  
 NID g1198225  
 VERSION N55346.1 GI:1198225  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 412)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B., Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Madis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Scheilenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 97044478

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.jnl.gov) for further information.  
 Insert Length: 517 Std Error: 0.00  
 Seq primer: m13 -40 forward  
 High quality sequence stop: 366.  
 Location/Qualifiers  
 1..412  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3795105"  
 /db\_xref="taxon:9606"  
 /clone\_image="245859"  
 /clone\_lib="Soares fetal liver spleen INFLS"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen. Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 123 a 86 c 92 g 111 t  
 ORIGIN

Query Match 7.7%; Score 17; DB 25; Length 412;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 TTGGACAAGATGATG 188  
 |||||||  
 Db 320 TTGGACAAGATGATG 336

RESULT 10  
 N77352/c 220 bp mRNA EST 01-FEB-1997  
 LOCUS N77352  
 DEFINITION yv47d02.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
 IMAGE:245859 5', mRNA sequence.  
 ACCESSION N77352  
 NID g1239930  
 VERSION N77352.1 GI:1239930  
 KEYWORDS EST.

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT  
FEATURES  
SOURCE  
BASE COUNT  
ORIGIN  
Query Match  
Best Local Similarity  
Matches  
QY  
Db  
RESULT  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 220)  
Haller,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapell,B.,  
Chissoe,S., Dietrich,N., Dubuque,T., Favell,A., Gish,W.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Plange,C., Rifkin,L.,  
Roehlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,  
Trevaskis,E., Underwood,K., Wohlmann,P., Waterson,R., Wilson,R.  
and Maria,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
On Nov 29, 1993 this sequence version replaced g1:430088.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 517 Std Error: 0.00  
Seq primer: reverse ET  
High quality sequence stop: 216.  
Location/Qualifiers  
1..220  
/organism="Homo sapiens"  
/db\_xref="GDB:3795105"  
/db\_xref="taxon:9606"  
/clone="IMAGE:245859"  
/clone\_1lb="Soares fetal liver spleen 1NFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was plimed with a Pac I - oligo(dT) primer  
15' AACGACGAGATTAATTAAGATCTTTTCTTTTCTTTT 3',  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."  
57 a 46 C 58 g 57 t 2 others  
7.7%; Score 17; DB 25; Length 220;  
100.0%; Pred. No. 41;  
0; Mismatches 0; Indels 0; Gaps 0;  
17: Conservative  
172 TTGGACAAAGATTGTC 188  
|||||  
120 TTGGACAAAGATTGTC 104  
W71854 260 bp mRNA EST 18-JUN-1996  
me45f05.r1 Soares mouse embryo NBMEL3.5 14.5 Mus musculus cDNA  
Clone IMAGE:390465 5', mRNA sequence.  
W71854  
g1381941  
W71854.1 GI:1381941  
EST.  
house mouse.  
Mus musculus  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE
AUTHORS
1 (bases 1 to 260)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubnue,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE
The Washu-HMT Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
On May 8, 1995 this sequence version replaced gi:800252.

Contact: Marra M/Mouse EST Project
Washu-HMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
WGI:242297
Seq primer: mob.REGA+ET.
Location/Qualifiers
1..260
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="390465"
/clone_id="Soares mouse embryo NDM613.5 14.5"
/sex="unknown"
/rna_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; site_1: Not I; site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer [5',
TGTTCACCAATCTGAAAGTGGAGCGCGCGAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]: double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pRT3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo. "
BASE COUNT
99 a 51 c 55 g 55 t
ORIGIN
Query Match 7.7%; Score 17; DB 26; Length 260;
Best Local Similarity 100.0%; Prid. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY 22 AACAAACGACACACAAA 38
|||||
Db 47 AACAAACGACACACAAA 63
|||||
RESULT 12
C06739/c C06739 292 bp mRNA EST 23-AUG-1996
DEFINITION C06739 Rat pancreatic islet cDNA Rattus norvegicus cDNA, mRNA
sequence.
ACCESSION C06739
NID G1503515
VERSION C06739.1 GI:1503515
KEYWORDS EST.
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 292)
Takeda,J.
Large scale collection of expressed sequence tags (ESTs) from rat

```

JOURNAL  
COMMENT  
pancreatic islet cDNA library  
Unpublished (1996)  
On Nov 29, 1993 this sequence version replaced g1:430322.

Contact: Jun Takeda  
Institute for Molecular and Cellular Regulation, Gunma University  
3-39-15 Shova-machi, Maebashi Gunma 371, Japan  
Tel: 272-20-8856  
Fax: 272-20-8896  
Email: jtakeda@sb.gunma-u.ac.jp.  
Location/Qualifiers

## FEATURES

Source

1..292  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/map="11"  
/clone\_lib="Rat pancreatic islet cDNA"  
/issue\_type="pancreatic islet"  
/note="Vector: Lambda ZAPri; Site\_1: EcoRI; Site\_2: XhoI;  
mRNA was prepared from normal rat islets. cDNA was  
directionally synthesized from the Xho I in the vector to  
the EcoRI site"  
BASE COUNT 66 a 85 c 60 g 76 t 5 others  
ORIGIN

Query Match 7.7%; Score 17; DB 27; Length 292;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 TGACACACAGCCACCA 146  
DB 34 TGACACACAGCCACCA 18  
|||||

RESULT 13  
LOCUS AA142568 213 bp mRNA EST 12-FEB-1997  
DEFINITION ms10c01.r1 Strataene mouse skin (#937313) Mus musculus cDNA clone  
IMAGE:606528 5', mRNA sequence.  
ACCESSION AA142568  
NID G1711831  
VERSION AA142568.1 GI:1711831  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
1 (bases 1 to 213)  
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE The WashU-HM Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On May 8, 1995 this sequence version replaced g1:801506.

Contact: Maria M/Mouse EST Project  
WashU-HM Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:371960  
Putative full length read  
vector to vector length is 214  
Seq primer: -28m13 rev1 ET from Amersham.

## FEATURES

Source

1..213  
/organism="Mus musculus"

/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:606528"  
/clone\_lib="Strataene mouse skin (#937313)"  
/sex="females"  
/issue\_type="whole skin"  
/dev\_stage="11 weeks old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: skin; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
0190 dt. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector: ~5'  
adaptor sequence: 5' GATTCCGACGAG 3' ~3' adaptor  
sequence: 5' CTCGAGTCTTTTCTTTTCTT 3' "  
BASE COUNT 57 a 31 c 37 g 88 t  
ORIGIN

Query Match 7.7%; Score 17; DB 29; Length 213;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 AACAAACACACACAA 38  
DB 146 AACAAACACACACAA 130  
|||||

RESULT 14  
LOCUS AA164101 313 bp mRNA EST 16-FEB-1997  
DEFINITION mt23507.r1 Soares mouse 3NBMS Mus musculus cDNA clone IMAGE:598332  
IMAGE:598332 5', mRNA sequence.  
ACCESSION AA164101  
NID G1740064  
VERSION AA164101.1 GI:1740064  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
1 (bases 1 to 313)  
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE The WashU-HM Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 25, 1995 this sequence version replaced g1:637882.

Contact: Maria M/Mouse EST Project  
WashU-HM Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:363764  
Seq primer: -28m13 rev2 from Amersham.

## FEATURES

Source

1..313  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:598332"  
/clone\_lib="Soares mouse 3NBMS"  
/sex="male"  
/issue\_type="Spleen"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT713D-Pac (Pharmacia) with a modified

polylinker: Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCCGCGCTGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library was constructed through three rounds of normalization, and was constructed by Bento Soares and M.Fátima Bonaldo."

```

Query Match      7.7%; Score 17; DB 29; Length 313;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 AACCAAAATCAAAAATG 67
    |||||||
Db 11 AACCAAAATCAAAAATG 27

```

RESULT	15
AAL6384/c	
LOCUS	
DEFINITION	AA166384.1 339 bp mRNA EST 19-DEC-1996 m7c3a06.t1 Soares mouse lymph node NBM LN Mus musculus cdna clone
ACCESSION	AA166384
NID	g1744041
VERSION	AA166384.1 GI:1744041
KEYWORDS	EST.
SOURCE	house mouse. Mus musculus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 339)
AUTHORS	Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubucque, Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thaising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The Mashu-HMMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced g1:1325261.

Contact: Marie M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel.: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact  
 IMAGE Consortium (info@image.llnl.gov) for further info.  
 MGI:387498  
 Seq primer: -28M13 rev2 from Amersham.  
 Location/Qualifiers  
 1..339

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image:635506"
/clone_lib="Scars mouse lymph node NBMN"
/sex="male"
/tissue_type="lymph node"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: lymph node; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site 1: Not I; Site 2: Eco RI
1st strand cDNA was primed with a Not I - oligo(dT) primer
5.
TGTACCAATCTGAAGTGGAGCGCCCGACATCTTTTTTTTTTTTTTTTTTT
3'; double-stranded cDNA was ligated to Eco RI adaptors

```

```

              (Pharmacia), digested with Not I and cloned into the Not
              and Eco RI sites of the modified pT73 vector. RNA
              provided by Dr. Bertrand Jordan. Library constructed and
              normalized by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      97 a      66 c      64 g      112 t
ORIGIN
Query Match      7.7%; Score 17; DB 29; Length 339;
Best Local Similarity 100.0%; Prid. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      22 AACAAACCACACACAAA 38
        |||||||
Db      242 AACAAACCACACACAAA 226

```

Search completed: August 6, 1999, 09:22:33  
Job time: 6224 sec

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 The Mashu-HHMI Mouse EST Project  
 Unpublished (1996)  
 On Nov 29, 1993 this sequence version replaced g1:430408.

Contact: Maria M/Mouse EST Project  
 Mashu-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:476156  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 377.

## FEATURES

source

Location/Qualifiers  
 1..429  
 /organism="Mus musculus"  
 /strain="B6D2 F1/3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:791748"  
 /clone\_1lb="Knowles Solter mouse 2 cell"  
 /tissue\_type="embryo"  
 /dev\_stage="2-cell"  
 /lab\_host="DH10B"  
 /note="Organ: embryo; Vector: pBluescribe (modified);  
 Site\_1: MluI; Site\_2: SalI; Cloned unidirectionally from  
 mRNA prepared from 13,500 2-cell stage embryos. Primer:  
 SalI(dT): 5'-CGGTCAGCCGTCGACGCTTTTCTTTT-3'. CDNAS  
 were cloned into the MluI/SalI sites of a modified  
 pBluescribe vector using commercial linkers (NEB).  
 Average insert size: 1.2 kb."

BASE COUNT 145 a 90 c 92 g 102 t  
 ORIGIN

Query Match 9.4%; Score 17; DB 33; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 AACAAACCAACCAACAA 151  
 ||||||||||||||||  
 DB 67 AACAAACCAACCAACAA 83

RESULT 15  
 AA488989 434 bp mRNA EST 15-AUG-1997  
 LOCUS aa54b11.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:824733 3',  
 DEFINITION mRNA sequence.  
 ACCESSION AA488989  
 NID g2218591  
 VERSION AA488989.1 GI:2218591  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 434)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced g1:1402313.

JOURNAL COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,

Ph.D., Gerald Marti, M.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/db/rp/image/image.html](http://www.bio.lnl.gov/db/rp/image/image.html)

Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 411.

## FEATURES

source

Location/Qualifiers  
 1..434  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="873p10:12"  
 /clone="IMAGE:824733"  
 /clone\_1lb="NCI\_CGAP\_GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAGTGGAGGGCGCCCTCATTTTTTTTTTTTTTTT-  
 3']. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pRT3D vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 119 a 73 c 92 g 150 t  
 ORIGIN

Query Match 9.4%; Score 17; DB 34; Length 434;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 ACAGAACAAACCAACAA 147  
 ||||||||||||||||  
 DB 82 ACAGAACAAACCAACAA 66

Search completed: August 6, 1999, 09:22:32  
 Job time: 6223 sec



```

DB      84  ATAAAGTACAGACTC 68

RESULT 10
LOCUS   AA142568/c      213 bp      mRNA      EST      12-FEB-1997
DEFINITION ms10c01.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:606528 5', mRNA sequence.
ACCESSION AA142568
NID       91711831
VERSION   AA142568.1  GI:1711831
KEYWORDS  EST
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 213)
           Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
           Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
           Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
           Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
           Waterston,R.
           The WashU-HMI Mouse EST Project
           Unpublished (1996)
           On May 8, 1995 this sequence version replaced gi:801506.

TITLE
JOURNAL
COMMENT

FEATURES
    source
        1..213
        /organism="Mus musculus"
        /strain="C57BL/6"
        /db_xref="taxon:10090"
        /clone="IMAGE:606528"
        /clone_1ib="Stratagene mouse skin (#937313)"
        /sex="females"
        /tissue_type="whole skin"
        /dev_stage="11 weeks old"
        /lab_host="SOLR (kanamycin resistant)"
        /note="Organ: skin; Vector: pBluescript SK-; Site_1:
        EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
        oligo dr. Whole skin from 11 week old C57BL/6 female mice.
        Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
        adaptor sequence: 5' GATTCGCGCAGAG 3' -3' adaptor
        sequence: 5' CTCGAGCTTTTGTTTTGTTTT 3'"

BASE COUNT      57 a      31 c      37 g      88 t

ORIGIN
Query Match      9.4%; Score 17; DB 29; Length 213;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      135  AACCAACACACACAAA 151
LOCUS   AA164101      313 bp      mRNA      EST      16-FEB-1997
DEFINITION ms10c01.r1 Soares mouse 3nDMS Mus musculus cDNA clone IMAGE:598332
IMAGE:598332 5', mRNA sequence.
ACCESSION AA164101
NID       91740064
VERSION   AA164101.1  GI:1740064
KEYWORDS  EST
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 313)
           Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
           Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
           Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
           Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
           Waterston,R.
           The WashU-HMI Mouse EST Project
           Unpublished (1996)
           On Jan 25, 1995 this sequence version replaced gi:637882.

TITLE
JOURNAL
COMMENT

FEATURES
    source
        1..313
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="IMAGE:598332"
        /clone_1ib="Soares mouse 3nDMS"
        /sex="male"
        /tissue_type="Spleen"
        /dev_stage="4 weeks"
        /lab_host="DH10B"
        /note="Vector: p773D-Pac (Pharmacia) with a modified
        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
        was primed with a Not I - oligo(dr) primer [5'
        TGTTTCATCTGAGTGGAGCGCCGCTGTTTGTTTTGTTTTGTTTT
        3']; double-stranded cDNA was ligated to Eco RI adaptors
        (Pharmacia), digested with Not I and cloned into the Not I
        and Eco RI sites of the modified p773 vector. RNA
        provided by Dr. Bertrand Jordan. Library went through
        three rounds of normalization, and was constructed by
        Bento Soares and M. Fatima Bonaldo."

BASE COUNT      134 a      56 c      52 g      71 t

ORIGIN
Query Match      9.4%; Score 17; DB 29; Length 313;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      164  AAGCAAAATCAAAAATG 180
LOCUS   AA16384/c      339 bp      mRNA      EST      19-DEC-1996
DEFINITION mt73a06.r1 Soares mouse lymph node NBMLN Mus musculus cDNA clone
IMAGE:635506 5', mRNA sequence.
ACCESSION AA16384
NID       91744041
VERSION   AA16384.1  GI:1744041

```

```

DEFINITION mt23g07.r1 Soares mouse 3nDMS Mus musculus cDNA clone IMAGE:598332
5', mRNA sequence.
ACCESSION AA164101
NID       91740064
VERSION   AA164101.1  GI:1740064
KEYWORDS  EST
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 313)
           Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
           Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
           Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
           Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
           Waterston,R.
           The WashU-HMI Mouse EST Project
           Unpublished (1996)
           On Jan 25, 1995 this sequence version replaced gi:637882.

TITLE
JOURNAL
COMMENT

FEATURES
    source
        1..313
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="IMAGE:598332"
        /clone_1ib="Soares mouse 3nDMS"
        /sex="male"
        /tissue_type="Spleen"
        /dev_stage="4 weeks"
        /lab_host="DH10B"
        /note="Vector: p773D-Pac (Pharmacia) with a modified
        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
        was primed with a Not I - oligo(dr) primer [5'
        TGTTTCATCTGAGTGGAGCGCCGCTGTTTGTTTTGTTTTGTTTT
        3']; double-stranded cDNA was ligated to Eco RI adaptors
        (Pharmacia), digested with Not I and cloned into the Not I
        and Eco RI sites of the modified p773 vector. RNA
        provided by Dr. Bertrand Jordan. Library went through
        three rounds of normalization, and was constructed by
        Bento Soares and M. Fatima Bonaldo."

BASE COUNT      134 a      56 c      52 g      71 t

ORIGIN
Query Match      9.4%; Score 17; DB 29; Length 313;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      164  AAGCAAAATCAAAAATG 180
LOCUS   AA16384/c      339 bp      mRNA      EST      19-DEC-1996
DEFINITION mt73a06.r1 Soares mouse lymph node NBMLN Mus musculus cDNA clone
IMAGE:635506 5', mRNA sequence.
ACCESSION AA16384
NID       91744041
VERSION   AA16384.1  GI:1744041

```

Query Match 9.4%; Score 17; DB 26; Length 260;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 135 AACAAACACACACAA 151  
|||||  
Db 47 AACAAACACACACAA 63

## RESULT 8

W80984/c 589 bp mRNA EST 17-OCT-1996  
LOCUS zdb4e06.r1 Soares\_fetal\_heart\_NBHL19W Homo sapiens CDNA clone  
DEFINITION IMAGE:347362 5', mRNA sequence.

ACCESSION W80984  
NID G1392085  
VERSION W80984.1 GI:1392085  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 589)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.

TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On May 8, 1995 this sequence version replaced gi:800146.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 642 Std Error: 0.00  
Seq primer: mod.RGA+RT  
High quality sequence stop: 366.  
Location/Qualifiers

## FEATURES

source

1. 589  
/organism="Homo sapiens"  
/db\_xref="GDB:1272737"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:347362"  
/clone\_lib="Soares\_fetal\_heart\_NBHL19W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - 01190(dt) primer (5'  
TGTTACCAATCTGAAGTGGAGCGCGCTCTTTTCTTTTCTTTT 3')  
double-stranded cDNA was size selected, ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NBHL19W."

BASE COUNT 185 a 104 c 116 g 181 t 3 others  
ORIGIN

Query Match 9.4%; Score 17; DB 26; Length 589;  
Best Local Similarity 100.0%; Pred. No. 42;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 37 TGAAGAGATGTGACT 53  
|||||  
Db 548 TGAAGAGATGTGACT 532

RESULT 9  
AA122783 524 bp mRNA EST 16-FEB-1997  
LOCUS mr02h12.r1 Soares mouse 3NBMS Mus musculus CDNA clone IMAGE:596327  
DEFINITION 5', mRNA sequence.

ACCESSION AA122783  
NID 91681740  
VERSION AA122783.1 GI:1681740  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 524)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 1, 1995 this sequence version replaced.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouse@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:361759  
Seq primer: -28M13 rev2 from Amerisham  
High quality sequence stop: 492.  
Location/Qualifiers

## FEATURES

source

1. 524  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/map="21"  
/clone\_image="IMAGE:596327"  
/clone\_lib="Soares mouse 3NBMS"  
/sex="male"  
/tissue\_type="Spleen"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - 01190(dt) primer (5'  
TGTTACCAATCTGAAGTGGAGCGCGCTCTTTTCTTTTCTTTT 3')  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. RNA  
provided by Dr. Bertrand Jordan. Library went through  
three rounds of normalization, and was constructed by  
Bento Soares and M.Fatima Bonaldo."

BASE COUNT 181 a 107 c 87 g 147 t 2 others  
ORIGIN

Query Match 9.4%; Score 17; DB 28; Length 524;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 18 ATAAAGTAACGAGACTC 34



```

/db_xref="taxon:9606"
/clone="IMAGE:2249902"
/clone_id="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and as
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
985608-986759, 1101192-1101959, and 121928-1220615)."
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT      187 a      67 c      85 g      186 t
ORIGIN

Query Match      9.4%; Score 18; DB 49; Length 525;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      160 AACACGCAAAATCAAA 177
      |||||||
Db      325 AACACGCAAAATCAAA 342

RESULT 6
LOCUS      M75917      443 bp      mRNA      EST      16-SEP-1992
DEFINITION CE06D10 Chris Martin sorted cDNA library Caenorhabditis elegans
            cDNA clone cm06d10 5', mRNA sequence.
ACCESSION      M75917
VERSION      G275240
KEYWORDS      M75917.1 GI:275240
SOURCE      EST.
ORGANISM      Caenorhabditis elegans.
REFERENCE      Caenorhabditis elegans.
AUTHORS      Eukaryota: Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditidae;
            Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
            1 (bases 1 to 443)
            Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A.,
            Hillier, L., Durbin, R.K., Green, P., Shownkeen, R., Halloran, N.,
            Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J.
            and Sulston, J.
            A survey of expressed genes in Caenorhabditis elegans
            Nature Genet. 1, 114-123 (1992)
TITLE      Nature Genet. 1, 114-123 (1992)
JOURNAL
MEDLINE
COMMENT      93250982

Contact: Waterston R.H. (USA) and Sulston J.E. (UK)
(USA) Dept. of Genetics or (UK)
(USA) Washington Univ. School of Medicine or (UK) MRC Laboratory of
Molecular Biology
Box 8232, 4566 Scott Ave., St. Louis, MI 63110, USA, or, Hills
Road, Cambridge CB2 2QH, UK
Tel: (USA) (314)3627072 or (UK) (0223)248011
Fax: (USA) (314)3624137 or (UK) (0223)402008
Email: r.henematode.wustl.edu or jesse@rc-lmba.cambridge.ac.uk
single read.

FEATURES
source      Location/Qualifiers
            1..443
            /organism="Caenorhabditis elegans"
            /strain="Bristol N2"
            /db_xref="taxon:6239"
            /clone="cm06d10"
            /clone_id="NCI_CGAP_Pr28"
            /lab_host="DH10B"
            /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker; Plasmid DNA from the
            normalized library NCI_CGAP_Pr22 was prepared, and as
            circles were made in vitro. Following HAP purification,
            this DNA was used as tracer in a subtractive hybridization
            reaction. The driver was PCR-amplified cDNAs from a pool
            of 5,000 clones made from the same library (clones
            985608-986759, 1101192-1101959, and 121928-1220615)."
            Subtraction by Bento Soares and M. Fatima Bonaldo.
            SHLX2 (Lipshitz, D.H. et al., Gene 88:25-36 (1990)) Host:

```

```

BASE COUNT      126 a      107 c      91 g      117 t      2 others
ORIGIN

Query Match      9.4%; Score 17; DB 20; Length 443;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      120 TTGAATCTGTACAGAA 136
      |||||||
Db      57 TTGAATCTGTACAGAA 41

RESULT 7
LOCUS      W71854      260 bp      mRNA      EST      18-JUN-1996
DEFINITION M645F05.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:390465 5', mRNA sequence.
ACCESSION      W71854
VERSION      G1381941
KEYWORDS      W71854.1 GI:1381941
SOURCE      EST.
ORGANISM      house mouse.
REFERENCE      Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 260)
            Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
            Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
            Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
            Thielsing, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
            Waterston, R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
            On May 8, 1995 this sequence version replaced gi:800252.

TITLE      The WashU-HMI Mouse EST Project
JOURNAL
COMMENT      On May 8, 1995 this sequence version replaced gi:800252.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouse@est.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:242297
Seq primer: mob.BEAG+ET.

FEATURES
source      Location/Qualifiers
            1..260
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:390465"
            /clone_id="Soares mouse embryo NbME13.5 14.5"
            /sex="unknown"
            /tissue="embryo"
            /dev_stage="13.5-14.5dpc total fetus"
            /lab_host="DH10B"
            /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker; Plasmid DNA from the
            normalized library NCI_CGAP_Pr22 was prepared, and as
            circles were made in vitro. Following HAP purification,
            this DNA was used as tracer in a subtractive hybridization
            reaction. The driver was PCR-amplified cDNAs from a pool
            of 5,000 clones made from the same library (clones
            985608-986759, 1101192-1101959, and 121928-1220615)."
            Subtraction by Bento Soares and M. Fatima Bonaldo.
            SHLX2 (Lipshitz, D.H. et al., Gene 88:25-36 (1990)) Host:

```

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:517447  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 377.  
 Location/Qualifiers

## FEATURES

## SOURCE

1.427  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /map="13"  
 /clone\_1db="IMAGE:889487"  
 /clone\_1db="Barstead mouse pooled organs MPLRB4"  
 /sex="mixed"  
 /tissue\_type="pooled organs"  
 /dev\_stage="7 day"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTCAGATCTGAGAGTGGAGCGCCGCTTTTCTTTTCTTTTCTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 (GTGATTCGCTACC), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 Library constructed by Bob Barstead."  
 BASE COUNT 143 a 108 c 83 g 93 t  
 ORIGIN

Query Match 9.9%; Score 18; DB 34; Length 427;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 119 GTTGAATCTGTACAGAA 136  
 ||||||||||||||||  
 Db 73 GTTGAATCTGTACAGAA 90

RESULT 4  
 AI309264 384 bp mRNA EST 01-FEB-1999  
 LOCUS g073e08.x1 NCI-CGAP\_Kids Homo sapiens cDNA clone IMAGE:1914182 3'  
 DEFINITION similar to gb:M96322 GRAVIN (HUMAN); mRNA sequence.  
 ACCESSION AI309264  
 NID 94004135  
 VERSION AI309264.1 GI:4004135  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 384)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Aug 21, 1998 this sequence version replaced.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www-bio.llnl.gov/bdrip/image/image.html

Trace considered overall poor quality  
 Insert Length: 438 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 1.  
 Location/Qualifiers

## FEATURES

## SOURCE

1.384  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="17q21:18q11.21"  
 /clone="IMAGE:1914182"  
 /clone\_1db="NCI-CGAP\_Kids"  
 /tissue\_type="2 pooled tumors (clear cell type)"  
 /lab\_host="DH10B"  
 /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 AACGGAAGATTCGCGCCGCAATATTTTCTTTTCTTTTCTTTT 3']  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 129 a 91 c 76 g 87 t  
 ORIGIN

Query Match 9.9%; Score 18; DB 44; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 35 TATGAAAAGGATGTGAG 52  
 ||||||||||||||||  
 Db 233 TATGAAAAGGATGTGAG 250

RESULT 5  
 A1659206 525 bp mRNA EST 10-MAY-1999  
 LOCUS tu02e12.x1 NCI-CGAP\_P28 Homo sapiens cDNA clone IMAGE:2249902 3'  
 DEFINITION similar to contains MER15.B1 MER15 repetitive element;; mRNA  
 sequence.  
 ACCESSION A1659206  
 NID 94762776  
 VERSION A1659206.1 GI:4762776  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 525)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2948292.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www-bio.llnl.gov/bdrip/image/image.html  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 459.  
 Location/Qualifiers  
 1.525  
 /organism="Homo sapiens"

```

VERSION      AAS81198.1  GI:2358970
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE    1 (bases 1 to 480)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      On May 9, 1995 this sequence version replaced gi:803001.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies Inc., David Krizman,
Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbfp/image/image.html

Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 405.
Location/Qualifiers
1. 480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1b="IMAGE:802572"
/clone_1ib="NCI-CGAP_CO1"
/tissue_type="Duk tumor"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT2; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT."

BASE COUNT   168 a 101 c 97 g 114 t

ORIGIN
Query Match      58.6%; Score 106; DB 35; Length 480;
Best Local Similarity 100.0%; Pred. No. 1.5e-45;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 GGAGAGCGCTTCTATATGTTTGCACACATGTTGATTCATAGTTCGTACAGA 135
    |||||||
DB 177 GGAGAGCGCTTCTATATGTTTGCACACATGTTGATTCATAGTTCGTACAGA 236
    |||||||

OY 136 ACAAAACACACACAAAGAGCTCCAAACAGCAAAATCAAAAATGC 181
    |||||||
DB 237 ACAAAACACACACAAAGAGCTCCAAACAGCAAAATCAAAAATGC 282
    |||||||

RESULT 2
AV026729 199 bp mRNA EST 02-JUN-1999
LOCUS     AV026729 Mus musculus adult C57BL/6J liver Mus musculus cDNA clone
DEFINITION 1300002D03, mRNA sequence.
ACCESSION AV026729
NID       94806339
VERSION   AV026729.1 GI:4806339
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 199)
AUTHORS    Carlinici, P., Shibata, K., Ozawa, Y., Konno, H., Ito, M., Aizawa, K.,
            Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
            Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
            Kawai, T., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H.,

```

```

Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,
Tomimaga, N., Watanabe, S., Yagame, H., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
On May 18, 1998 this sequence version replaced gi:3138591.

Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rc.riken.go.jp
Thermotabilization and thermostabilization of thermostable enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1. 199
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="18; Chromosomes 7,11,12; 21q"
/clone_1b="1300002D03"
/clone_1ib="Mus musculus adult C57BL/6J liver"
/tissue_type="liver"
/dev_stage="adult"

BASE COUNT   56 a 24 c 44 g 74 t 1 others

ORIGIN
Query Match      10.5%; Score 19; DB 50; Length 199;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 133 AGACAAACACACACAA 151
    |||||||
DB 82 AGACAAACACACACAA 64
    |||||||

RESULT 3
AA498535 427 bp mRNA EST 01-JUL-1997
LOCUS     vha0f12.r1 Barstead mouse pooled organs MPRB4 Mus musculus cDNA
DEFINITION CLONE IMAGE:889487 5' similar to TR:G1184066 G1184066
ACCESSION AA498535
NID       92233558
VERSION   AA498535.1 GI:2233558
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 427)
AUTHORS    Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
            Geisel, S., Kuwada, T., Lacy, M., Le, M., Martin, J., Morris, M.,
            Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
            Theisling, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
            Waterston, R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
            On Sep 12, 1996 this sequence version replaced gi:1294625.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

```

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:28 ; Search time 3059.95 Seconds  
(without alignments)  
116.678 Million cell updates/sec

Title: US-09-049-696-4  
Perfect score: 181  
Sequence: 1 CAAGAAGTGCACATTCAATTA.....ACAGCAAAATCAAAATGTC 181

Scoring table: OLIGO\_NUC

Searched: 2546578 segs, 98626752 residues

Database : EST.\*

1: em\_est1:.\*  
2: em\_est2:.\*  
3: em\_est3:.\*  
4: em\_est4:.\*  
5: em\_est5:.\*  
6: em\_est6:.\*  
7: em\_est7:.\*  
8: em\_est8:.\*  
9: em\_est9:.\*  
10: em\_est10:.\*  
11: em\_est11:.\*  
12: em\_est12:.\*  
13: em\_est13:.\*  
14: em\_est14:.\*  
15: em\_est15:.\*  
16: em\_est16:.\*  
17: em\_est17:.\*  
18: em\_est18:.\*  
19: em\_est19:.\*  
20: em\_est20:.\*  
21: em\_est21:.\*  
22: em\_est22:.\*  
23: em\_est23:.\*  
24: em\_est24:.\*  
25: em\_est25:.\*  
26: em\_est26:.\*  
27: em\_est27:.\*  
28: em\_est28:.\*  
29: em\_est29:.\*  
30: em\_est30:.\*  
31: em\_est31:.\*  
32: em\_est32:.\*  
33: em\_est33:.\*  
34: em\_est34:.\*  
35: em\_est35:.\*  
36: em\_est36:.\*  
37: em\_est37:.\*  
38: em\_est38:.\*  
39: em\_est39:.\*  
40: em\_est40:.\*  
41: em\_est41:.\*  
42: em\_est42:.\*  
43: em\_est43:.\*  
44: em\_est44:.\*  
45: em\_est45:.\*  
46: em\_est46:.\*  
47: em\_est47:.\*  
48: em\_est48:.\*  
49: em\_est49:.\*  
50: em\_est50:.\*  
51: em\_est51:.\*  
52: em\_est52:.\*  
53: em\_est53:.\*

54: em\_est22:.\*  
55: em\_est23:.\*  
56: em\_est24:.\*  
57: em\_est25:.\*  
58: em\_est26:.\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	106	58.6	480	35	AA581198	AA581198 nd38c07.r
2	19	10.5	199	50	AV026729	AV026729
3	18	9.9	427	34	AA498535	AA498535 vH40f12.r
4	18	9.9	384	44	AI309264	AI309264 q073e08.x
5	18	9.9	525	49	AI659206	AI659206 tu02e12.x
6	17	9.4	443	20	M75917	M75917 CE06010.Ch
7	17	9.4	260	26	W71854	W71854 me45f05.r1
8	17	9.4	589	26	W80984	W80984 zd84e06.r1
9	17	9.4	524	28	AA122783	AA122783 mr02h12.r
10	17	9.4	213	29	AA142568	AA142568 ms10c01.r
11	17	9.4	313	29	AA164101	AA164101 mr23g07.r
12	17	9.4	339	29	AA166384	AA166384 mt73a06.r
13	17	9.4	361	29	AA178674	AA178674 mt18h03.r
14	17	9.4	429	33	AA414677	AA414677 vd06e07.s
15	17	9.4	434	34	AA488989	AA488989 aa54b11.s
16	17	9.4	476	36	AA636732	AA636732 vr20d08.r
17	17	9.4	525	37	AA696398	AA696398 GM07607.s
18	17	9.4	346	37	AA704201	AA704201 zj77e10.s
19	17	9.4	476	37	AA734960	AA734960 v578d08.r
20	17	9.4	372	38	AA744665	AA744665 ny79h11.s
21	17	9.4	463	38	AA809480	AA809480 ob85c03.s
22	17	9.4	379	39	AA813879	AA813879 ny95a06.s
23	17	9.4	472	39	AA815232	AA815232 oc05d10.s
24	17	9.4	363	40	AA871023	AA871023 op68f08.s
25	17	9.4	432	40	AA876402	AA876402 qg68f08.s
26	17	9.4	485	41	AI056234	AI056234 oy90f06.x
27	17	9.4	1051	41	AI068506	AI068506 m9ae0002d
28	17	9.4	534	42	AI083081	AI083081 MS737.Med
29	17	9.4	316	42	AI083085	AI083085 MS443J.Me
30	17	9.4	548	44	AA043085	AA043085 AA043085
31	17	9.4	386	46	AI451099	AI451099 mt18h03.x
32	17	9.4	313	48	AI592108	AI592108 mr23g07.y
33	17	9.4	388	48	AI592326	AI592326 mt18h03.y
34	17	9.4	454	48	AI604060	AI604060 vr20d08.y
35	17	9.4	412	50	AI674736	AI674736 wd19g12.x
36	17	9.4	652	51	AI725132	AI725132 1031.PtIF
37	17	9.4	481	51	AI733457	AI733457 OG68f08.x
38	16	8.8	254	21	RI3614	RI3614 yf63h06.r1
39	16	8.8	509	21	RI9757	RI9757 y940f02.r1
40	16	8.8	346	22	HI1862	HI1862 ym15e04.r1
41	16	8.8	490	22	HI4405	HI4405 ym18c08.r1
42	16	8.8	474	22	HI4489	HI4489 ym18c11.r1
43	16	8.8	380	23	R87634	R87634 yp89a07.s1
44	16	8.8	300	24	D65089	D65089 CELK130F7R
45	16	8.8	437	24	H73017	H73017 ys10a01.r1

## ALIGNMENTS

RESULT 1  
LOCUS AA581198  
DEFINITION nd38c07.r1 NCI CGAP Col Homo sapiens CDNA clone IMAGE:802572.5'  
similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.  
ACCESSION AA581198  
NID g2358970

EST 05-JAN-1998

Query Match 6.3%; Score 16; DB 1; Length 2216;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGCTTTGGGGCCCTT 149  
 |||||||  
 DB 2131 TGCTTTGGGGCCCTT 2116

## RESULT 14

075927/c  
 ID 075927 standard; DNA: 2272 BP.

AC 075927;  
 DT 17-AUG-1995 (first entry)  
 DE Mouse delta opioid receptor MOR1 cDNA.  
 KW Mouse; kappa; delta; mu; opioid receptor; brain; primer; PCR; amplify;  
 KW transmembrane domain; somatostatin; receptor; human; expression vector;  
 KW truncate; chimaeric; assay; probe; ss.  
 OS Mus musculus.

FN Key Location/Qualifiers  
 FT cds 12..1130  
 FT /tag= a  
 FT /product= mouse delta opioid receptor

PN MO928132-A.  
 PD 08-DEC-1994.  
 PF 20-MAY-1994; U05747.  
 PR 20-MAY-1993; US-066296.  
 PR 30-JUL-1993; US-100694.  
 PR 05-NOV-1993; US-147592.  
 PA (ARCH-) ARCH DEV CORP.  
 PI Bell GI, Reisine T, Yasuda K;  
 DR WPI: 95-022804/03.  
 DR P-PSDB: R67670.

PT polynucleotides and peptides derived from opioid receptor  
 PT screening assays for useful drug substances.

PS Clam 6; Page 215-221; 300pp; English.  
 CC The nucleotide sequence of the novel mouse delta opioid receptor gene  
 CC MOR1. The gene was isolated from a mouse brain cDNA library using a  
 CC fragment (amplified from the cDNA library with primers Q75929-30) as a  
 CC probe. The primers are based on the conserved sequences present in the  
 CC second and third transmembrane domains of somatostatin (SSTR) receptor  
 CC subtypes SSTR1, SSTR2 and SSTR3. The 1.3 kb EcoRT-SacI fragment from the  
 CC mouse delta opioid receptor clone, lambda ms1-2, was subcloned into the  
 CC CMV promoter-based expression vector pCMV-6c. The resultant construct  
 CC pCMV-ms1-2 was transfected into COS-1 cells for protein production. The  
 CC gene encoding the opioid receptor can be used to produce complete,  
 CC truncated or chimaeric opioid receptor proteins. The opioid receptors  
 CC thus produced are useful for the development of novel assays designed to  
 CC select or improve substances, capable of interacting with the opioid  
 CC receptor proteins, for use in diagnosis, drug design and therapeutic  
 CC applications.  
 SQ Sequence 2272 BP; 485 A; 665 C; 650 G; 472 T;

Query Match 6.3%; Score 16; DB 1; Length 2272;  
 Best Local Similarity 100.0%; Pred. No. 28;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGCTTTGGGGCCCTT 149  
 |||||||  
 DB 2073 TGCTTTGGGGCCCTT 2058

## RESULT 15

T02795  
 ID T02795 standard; CDNA to mRNA; 2911 BP.

DT 24-APR-1996 (first entry)  
 DE Human glutamate receptor subunit cDNA (GluRAfl1pG).  
 KW glutamate receptor subunit; human; flip; splice variant;  
 KW identify ligand; treatment; ischaemia; schizophrenia; epilepsy; ss.

OS Homo sapiens.  
 FN Key Location/Qualifiers  
 FT cds 144..2864  
 FT /tag= a  
 FT /product= glutamate receptor subunit

PN MO9521188-A1.  
 PD 10-AUG-1995.  
 PF 27-JAN-1995; E00290.  
 PR 07-FEB-1994; DE-403666.  
 PA (BADI ) BASF AG.  
 PI Bach A, Hoeger T, Lemaire H, Sterrer S, Ullsch A;  
 DR WPI: 95-283726/37.  
 DR P-PSDB: R84912.

PT DNA sequences coding for glutamate receptor sub-unit variants -  
 PT useful for the identification of functional ligands for these  
 PT receptors and for preparing anti-sense reagents  
 PS Clam 1; Page 13-18; 71pp; English.

CC T02795-802 are DNA sequences encoding variants of the human glutamate  
 CC receptor (GluR) subunits. cDNAs encoding GluR subunits A, B, C, and D,  
 CC were isolated from the brain. There are 2 splice variants, "flip" and  
 CC "flop", for each subunit that are known. It was found that for GluRA,  
 CC GluRB, GluRC and GluRD, the first amino acid of the "flip/flop"-region,  
 CC can be Glycine (G) or Arginine (R), through different RNA splicing.  
 CC For the GluRB subunit, it was found that it is possible to have a  
 CC glutamine (Q) or an R in the transmembrane domain, before the flip/flop  
 CC region giving a total of 8 possible conformations. The GluRAfl1pR cDNA  
 CC possesses in comparison to GluRAfl1pG cDNA, a base change at position  
 CC 2269, a glycine codon (GGA) is converted to an arginine codon (AGA)  
 CC (sic). The same applies for GluRAfl1pR. Cells transfected with the DNA  
 CC are useful for identifying functional ligands for GluR. Glutamate is  
 CC an important excitatory neurotransmitter in the central nervous system  
 CC and is involved in the pathophysiology of e.g., epilepsy, schizophrenia  
 CC and ischaemia. GluR ligands have the potential to be used in the  
 CC treatment of these conditions.  
 SQ Sequence 2911 BP; 775 A; 674 C; 799 G; 663 T;

Query Match 6.3%; Score 16; DB 1; Length 2911;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 ATGATGGCAGTGA 243  
 |||||||  
 DB 1831 ATGATGGCAGTGA 1846

Search completed: August 5, 1999, 17:40:29  
 Job time: 6264 sec



ID	Accession	Score	Length	Start	End	Strand	Seq	Ref	Match	Mismatch	Indel	Gap
PT	New polypeptide having binding affinity to inositol-3-phosphate -											
PT	prepd. by culturing cell contg. recombinant plasmid comprising											
PT	DNA and vector DNA											
PS	Disclosure: Fig 2(1-3); 11pp; Japanese.											
CC	The sequence may be included in a plasmid/vector for transformation											
CC	of a host cell and mass-prodn. of the peptide.											
CC	Sequence 9871 BP; 2687 A; 2325 C; 2555 G; 2304 T;											
SO												
Query Match		6.3%	Score 16;	DB 1;	Length 9871;							
Best Local Similarity		100.0%	Pred. No. 27;									
Matches 16;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;				
OY	208 CCCTCCAGACAGCCA 223											
DB	2432 CCCTCCAGACAGCCA 2417											
RESULT 10												
ID	Q25975	standard; DNA; 15377 BP.										
AC	Q25975;											
DT	08-JAN-1993 (first entry)											
DE	MH mutant porcine ryanodine receptor cDNA.											
KW	MH; RYR1; calcium release channel; sarcoplasmic reticulum;											
KW	transverse tubule; Pletetrain; Yorkshire; polymorphism; beta strand; ss.											
OS	Synthetic.											
PH	Key											
FT	cds	location/Qualifiers										
FT		130..15237										
FT	variation	/*tag= a										
FT		207										
FT	variation	/*tag= b										
FT		/label= Polymorphic_site										
FT	variation	405										
FT		/*tag= c										
FT		/label= Polymorphic_site										
FT	variation	438										
FT		/*tag= d										
FT	variation	/label= Polymorphic_site										
FT		876										
FT		/*tag= e										
FT	variation	/label= Polymorphic_site										
FT		1329										
FT	variation	/*tag= f										
FT		/label= Polymorphic_site										
FT	variation	1972										
FT		/*tag= g										
FT	variation	/label= MH_mutation										
FT		2007										
FT	variation	/*tag= h										
FT		/label= Polymorphic_site										
FT	variation	4071			</							

FT	/tag- p
FT	/label- Polymorphic-site
FT	9585
FT	/tag- q
FT	/label- Polymorphic-site
FT	9600
FT	/tag- r
FT	/label- Polymorphic-site
FT	9951
FT	/tag- s
FT	/label- Polymorphic-site
FT	10111
FT	/tag- t
FT	/label- Polymorphic-site
FT	11250
FT	/tag- u
FT	/label- Polymorphic-site
FT	12300
FT	/tag- v
FT	/label- Polymorphic-site
FT	14007
FT	/tag- w
FT	/label- Polymorphic-site
FT	15355 .15360
FT	/tag- x
PD	MO9211387-A.
PD	09-JUL-1992.
PF	20-DEC-1991; CA0457.
PR	21-DEC-1990; GB-027869.
PR	20-MAY-1991; GB-010865.
PR	09-SEP-1991; GB-019250.
PA	(UYGU-) UNIV GUELPH.
PA	(UTOR-) UNIV TORONTO INNOVATIONS FOUND.
DR	MacLennan DH, O'Brien PJ;
DR	WPI. 92-250106/30.
DR	P-P5DB; R25450.
PT	Purified DNA mol. for diagnosis of porcine malignant hyperthermia
PT	- comprises DNA sequence encoding normal or mutant ryanodine
PT	receptor with specified endonuclease restriction map
PS	Disclosure; Fig 2; 96pp; English.
CC	The sequence given is the mutant pig ryanodine receptor (RYR1) gene
CC	from swine cDNA. The polymorphic sites were observed in comparisons
CC	of Pietrain and Yorkshire breeds. There are 17 polymorphisms between
CC	the two breeds. The polymorphism at position 1972 causes a mutation
CC	from Arg to Cys and this is thought to be the molecular basis of
CC	porcine malignant hyperthermia (MH). This mutation lies within the
CC	region of RYR1 that is concerned with the binding of regulators of Ca <sup>2+</sup>
CC	release channel gating. Analysis of surrounding sequences suggests
CC	that this mutation lies within a beta strand domain comprising roughly
CC	of amino acids 520 to 830. RYR1 is the calcium release channel of the
CC	sarcoplasmic reticulum and is a large protein which spans the gap
CC	between the transverse tubule and the sarcoplasmic reticulum. The
CC	cannel is activated by ATP, calcium, caffeine, and micro-molar
CC	ryanodine. It is inhibited by ruthenium red, tetracaine, calmodulin,
CC	high Mg <sup>2+</sup> and ryanodine.
SQ	Sequence 15377 BP; 3197 A; 4630 C; 4755 G; 2774 T;
Query Match 6.3%; Score 16; DB 1; Length 15377;	
Best Local Similarity 100.0%; Pred. No. 27;	
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	127 TCATGATGCCTTTGG 142 
Db	14930 TCATTGATCCTTTGG 14945
RESULT 11	
ID	OS1025 standard; cDNA; 2718 BP.
AC	OS1025;
DE	18-APR-1994 (first entry)
DT	Human glutamate receptor gene, HSGU1r1.

DE glycol dehydrogenase domains I, II, III and IV.  
KW Enzyme; catechol; benzene; ds.  
OS Pseudomonas.  
FH Key Location/Qualifiers  
FT cds 617..1949  
FT /tag= a  
FT /product-domain I  
FT cds 2069..2633  
FT /tag= b  
FT /product-domain II  
FT cds 2963..4192  
FT /tag= c  
FT /product-domain III  
FT cds 4519..5016  
FT /tag= d  
FT /product-domain IV  
PN J6228276-A.  
PD 07-OCT-1987.  
PF 28-MAR-1986; 068700.  
PR 28-MAR-1986; JP-068700.  
PA (IDEX) Idemitsu Kosan KK.  
WP1: 87-323403/46.  
PT gene contg. domains that code aminoacid sequence - used for  
construction of gene suitable for transformation of prokaryotic cells.  
CC disclosure; fig 1, pp 361-2; 11pp; Japanese.  
CC The gene is used for construction of a gene suitable for  
transformation of prokaryotic cells (e.g. E.coli), and by using the  
cells, catechol or deriv. can be produced from a benzene cpd.  
SQ Sequence 5274 BP; 1057 A; 1493 C; 1673 G; 1051 T;

Query Match 6.3%; Score 16; DB 1; Length 5274;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GTGGTGCATCATCCA 22  
Db 3252 GTGGTGCATCATCCA 3237

RESULT 7  
ID N70822/c  
AC N70822; standard; DNA; 1230 BP.  
DT 16-JAN-1991 (first entry)  
DE DNA encoding enzyme protein.  
KW Enzyme protein; ds DNA; beta-tyrosinase gene; catechol gene; benzene;  
KW L-DOPA; Parkinson's disease.  
PN J62259590-A.  
PD 11-NOV-1987.  
PF 07-MAY-1986; J03196.  
PR 07-MAY-1986; JP-103196.  
PA (IDEX) Idemitsu Kosan KK.  
WP1: 87-358157/51.  
PT Complex recombinant DNA - contains DNA sequence for L-dopa prodn. from  
benzene, and beta-tyrosinase gene.  
PS Claim 1; page 502-3; 20pp; Japanese.  
CC The DNA (region III) forms part of a complex recombinant DNA which also  
contains gene regions I, II and IV and a beta-tyrosinase gene.  
CC The DNA encodes an enzyme protein. Regions I - IV make up a catechol  
producing gene derived from bacteria such as Pseudomonas. The beta  
tyrosinase gene is derived from bacteria such as Erwinia herbicola and  
Aerobacter aerogenes. Bacteria transformed with the recombinant  
DNA are capable of producing catechol from benzene, and have beta-  
tyrosinase activity. Using the transformant, L-dopa can be produced  
from benzene cheaply and safely. L-dopa is used as a remedy for  
Parkinson's disease.  
SQ Sequence 1230 BP; 210 A; 356 C; 426 G; 238 T;

Query Match 6.3%; Score 16; DB 1; Length 1230;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GTGGTGCATCATCCA 22  
Db 290 GTGGTGCATCATCCA 275

RESULT 8  
ID N70824/c  
AC N70824; standard; DNA; 5274 BP.  
DT 16-JAN-1991 (first entry)  
DE Catechol producing gene.  
KW Catechol gene; ds DNA; beta-tyrosinase gene; benzene; L-dopa;  
KW Parkinson's disease.  
FH Key Location/Qualifiers  
FT misc\_feature 617..1963  
FT /tag= a  
FT /note="Region I"  
FT misc\_feature 2071..2633  
FT /tag= b  
FT /note="Region II"  
FT misc\_feature 2963..4192  
FT /tag= c  
FT /note="Region III"  
FT misc\_feature 4519..5016  
FT /tag= d  
FT /note="Region IV"  
PN J62259590-A.  
PD 11-NOV-1987.  
PF 07-MAY-1986; J03196.  
PR 07-MAY-1986; JP-103196.  
PA (IDEX) Idemitsu Kosan KK.  
WP1: 87-358157/51.  
PT Complex recombinant DNA - contains DNA sequence for L-dopa prodn. from  
benzene, and beta-tyrosinase gene.  
PS disclosure; fig. 1; 20pp; Japanese.  
CC The catechol gene contains gene regions I, II, III and IV.  
CC The gene may form part of a complex recombinant DNA also contg. the  
beta-tyrosinase gene derived from, eg Erwinia herbicola or Aerobacter  
aerogenes. Bacteria transformed with the recombinant DNA are capable of  
producing catechol from benzene, and have beta-tyrosinase activity.  
CC Using the transformant, L-dopa can be produced from benzene cheaply and  
safely. L-dopa is used as a remedy for Parkinson's disease.  
SQ Sequence 5274 BP; 1054 A; 1491 C; 1673 G; 1056 T;

Query Match 6.3%; Score 16; DB 1; Length 5274;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GTGGTGCATCATCCA 22  
Db 3252 GTGGTGCATCATCCA 3237

RESULT 9  
ID Q13593/c  
AC Q13593; standard; DNA; 9871 BP.  
DT 27-NOV-1991 (first entry)  
DE Inositol-3-phosphate binding peptide.  
KW I-3-P; transformation; inositol; ss.  
FH Key Location/Qualifiers  
FT cds 329..8575  
FT /tag= a  
FT /label= I-3-P\_binding\_peptide  
PN J03183482-A.  
PD 09-AUG-1991.  
PF 14-DEC-1989; 324256.  
PR 14-DEC-1989; JP-324256.  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
WP1: 91-277584/38.  
DR P-PSDB; R13887.



DR WPI; 93-272882/34.  
 PT Enriched oligonucleotides and corresp. sequences - used as  
 PT markers for human genes transcribed in-vivo, facilitate tagging  
 PT of most human genes  
 PS Example 4: Page 418: 500pp: English.  
 CC The Expressed Sequence Tag was isolated from a human brain cDNA  
 CC library as part of a large set of ESTs which can be used as markers  
 CC for human genes transcribed in vivo. They can be used to facilitate  
 CC tagging of most human genes, for mapping locations of expressed genes  
 CC on chromosomes, for individual or forensic identification, for mapping  
 CC locations of disease-associated genes, for identification of tissue  
 CC type and for prep. of antisense sequences, probes and constructs.  
 CC EST01042 has a "poor" coding probability as evaluated using the  
 CC coding-region prediction program CRM. See also Q59041-Q61440.  
 SQ Sequence 314 BP; 72 A; 79 C; 57 G; 106 T;

Query Match 6.7%; Score 17; DB 1; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 AGACTAGAGGAGCTGT 69  
 Db 178 AGACTAGAGGAGCTGT 194

RESULT 3  
 ID V45211 standard: cDNA: 1199 BP.  
 AC V45211;  
 DT 19-OCT-1998 (first entry)  
 DE Human endosulfine B gene.  
 KW ds: endosulfine; central nervous system; pancreatic beta-cell; pain;  
 KM Insulin; cytoprotective agent; diabetes; epilepsy; depression; ischemia.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 107..472  
 FT /\*tag= a  
 FT /product= "Endosulfine B protein"  
 PN M09830692-A2.  
 PD 16-JUL-1998.  
 PE 07-JAN-1998: U00137.  
 PR 07-JAN-1997: US-779775.  
 PA (ABBO ) ABBOTT LAB.  
 PI Anderson KL, Roch J, Scott VE, Sullivan JP;  
 DR WPI; 98-399140/34.  
 P-PSDB: W61552.  
 PT Novel human endosulfine A and B - for screening for compounds  
 PT modulating endosulfine receptor activity, useful as cytoprotective  
 PT agent(s)  
 PS Claim 2: Page 58; 76pp: English.  
 CC The human endosulfine A and B genes encode proteins related to other  
 CC members of the endosulfine family which have been shown to bind receptors  
 CC from both the central nervous system and pancreatic beta-cells and to  
 CC induce insulin secretion from a rodent beta-cell tumour line. The  
 CC protein can be used to screen for compounds that modulate endosulfine  
 CC receptor activity, especially as a cytoprotective agent.  
 CC Modulators of endosulfine activity may represent a treatment of  
 CC diabetes, epilepsy, pain, depression and ischemia.  
 SQ Sequence 1199 BP; 335 A; 288 C; 333 G; 243 T;

Query Match 6.7%; Score 17; DB 1; Length 1199;  
 Best Local Similarity 100.0%; Pred. No. 8.5;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 AGACTAGAGGAGCTGT 69  
 Db 998 AGACTAGAGGAGCTGT 982

RESULT 4  
 N70677/c

ID N70677 standard: DNA; 1230 BP.  
 AC N70677;  
 DT 20-JAN-1991 (first entry)  
 DE Sequence encoding Pseudomonas sp. benzene oxygenase and benzene  
 DE glycol dehydrogenase domain III.  
 KM Enzyme; catechol; benzene; ds.  
 OS Pseudomonas.  
 PN J62282/6-A.  
 PD 07-OCT-1987.  
 PE 28-MAR-1986: 068700.  
 PR 28-MAR-1986: JP-068700.  
 PA (IDEX) Idemitsu Kosan Kk.  
 DR WPI; 87-323403/46.  
 PT Gene contg. domains that code aminoacid sequence - used for  
 PT construction of gene suitable for transformation of prokaryotic cells.  
 PS Claim 2: Page 354-5; 11pp: Japanese.  
 CC The gene is used for construction of a gene suitable for  
 CC transformation of prokaryotic cells (e.g. E.coli), and by using the  
 CC cells, catechol or deriv. can be produced from a benzene cpd.  
 SQ Sequence 1230 BP; 210 A; 360 C; 422 G; 238 T;

Query Match 6.3%; Score 16; DB 1; Length 1230;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GTGGTGCATCATCA 22  
 Db 290 GTGGTGCATCATCA 275

RESULT 5  
 ID N70053/c  
 AC N70053 standard: DNA; 1230 BP.  
 DT 03-MAY-1991 (first entry)  
 DE Benzene oxygenase segment III.  
 KM Benzene oxygenase; segment; plastic; aromatic cpd.; ds.  
 OS Escherichia coli SI-300 FERM BP-1358.  
 PN EP-247600-A.  
 PD 02-DEC-1987.  
 PE 27-MAY-1987: 107740.  
 PR 28-MAY-1986: JP-121263.  
 PA (IDEX) IDEMITSU KOSAN KK.  
 PI Doi S, Irie S;  
 DR WPI; 87-336271/48.  
 PT DNA sequence encoding polypeptide having benzene oxygenase  
 PT activity - used in prokaryotic cells to convert aromatic cpds. to  
 PT 1,2-di:hydroxy-cyclo:hexa-3,5-diene  
 PS Claim 1; Page 13-16; 20pp: English.  
 CC DNA sequences contg. DNA segments I(N70051), II(N70052) and III in  
 CC combination encode a polypeptide having a benzene oxygenase activity.  
 CC Prokaryotic cells transformed with these segments can be used to  
 CC convert aromatic cpds. in high yield to 1,2-dihydroxy-cyclohexa-3,5-  
 CC diene which can be used in the prodn. of engineering plastics.  
 SQ Sequence 1230 BP; 210 A; 357 C; 425 G; 238 T;

Query Match 6.3%; Score 16; DB 1; Length 1230;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GTGGTGCATCATCA 22  
 Db 290 GTGGTGCATCATCA 275

RESULT 6  
 N70679/c  
 ID N70679 standard: DNA; 5274 BP.  
 AC N70679;  
 DT 19-MAR-1991 (first entry)  
 DE Sequence encoding Pseudomonas sp. benzene oxygenase and benzene

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:40:25 ; Search time 650.92 seconds

(without alignments)  
97.245 Million cell updates/sec

Title: US-09-049-696-8

Perfect score: 253

Sequence: 1 AACAAAGTGTGGCATCATC.....GGCAGAGTGAATGTGACAG 253

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	7.1	1386	1	Rat pheromone rece
2	17	6.7	314	1	Human brain expres
3	17	6.7	1199	1	Human endosulfine
4	16	6.3	1230	1	Sequence encoding
5	16	6.3	1230	1	Benzene oxygenase
6	16	6.3	5274	1	Sequence encoding
7	16	6.3	1230	1	DNA encoding enzym
8	16	6.3	5274	1	Catechol producing
9	16	6.3	9871	1	Inositol-3-phospha
10	16	6.3	15377	1	MH mutant porcine
11	16	6.3	2718	1	Human glutamate re
12	16	6.3	3220	1	Human glurib recep
13	16	6.3	2216	1	Mutine delta opioi
14	16	6.3	2272	1	Mouse delta opioi
15	16	6.3	2911	1	Human glutamate re
16	16	6.3	2946	1	Human glutamate re
17	16	6.3	2218	1	Mouse delta opiate
18	16	6.3	2981	1	Human ehb21 encodi
19	16	6.3	1304	1	H. pylori GHPD 83
20	16	6.3	1837	1	H. pylori GHPD 84
21	16	6.3	44453	1	Human kidney amino
22	15	5.9	3905	1	Sequence encoding
23	15	5.9	3697	1	Human Natriuretic
24	15	5.9	888	1	Mouse kidney cell
25	15	5.9	3905	1	Human Factor XIII
26	15	5.9	466	1	Sequence of intron
27	15	5.9	5931	1	Sequence of the hu
28	15	5.9	14255	1	ALU-1 gene cDNA. D
29	15	5.9	1420	1	11q23 chromosome t
30	15	5.9	4201	1	CDNA encoding htrx
31	15	5.9	11907	1	CDNA encoding htrx
32	15	5.9	1455	1	ALU-1 (acute lymph
33	15	5.9	1556	1	Rat hippocampal ne
34	15	5.9	1200	1	MLL 1.5EB fragment
35	15	5.9	1420	1	MLL cDNA clone 14-
36	15	5.9	4201	1	Human Serrate-2 (H
37	15	5.9	4483	1	Human Serrate-2 (H
38	15	5.9	669	1	LsrDNA sequence fr
39	15	5.9	669	1	LsrDNA sequence fr
40	15	5.9	669	1	LsrDNA sequence fr
41	15	5.9	669	1	LsrDNA sequence fr
42	15	5.9	669	1	LsrDNA sequence fr
43	15	5.9	669	1	LsrDNA sequence fr

## ALIGNMENTS

```
RESULT 1
ID T69545 standard; DNA; 1386 BP.
AC T69545;
DE 26-AUG-1997 (first entry)
KW Pheromone receptor VNI cDNA.
KW Vomeronasal sensory neuron; social behaviour;
KW maternal behaviour; reproductive behaviour; fertility;
KW hormone secretion; ss.
OS Rattus sp.
FH Key Location/Qualifiers
FT Cds 270..1217
FT /*tag- a
FT
FT MO9714790-A1.
PD 24-APR-1997. U16637.
PE 18-OCT-1996: US-005698.
PR 19-OCT-1995: US-005698.
PA (UYCO ) UNIV COLDMBIA NEW YORK.
PI Axel R, Dulac C;
DR WPI; 97-245107/22.
DR P-PSDB; W19103.
PT Nucleic acid molecule encoding vertebrate pheromone receptor -
PT useful to identify modulators for control of reproductive and social
PT behaviour, fertility and hormone secretion
PS Claim 1; Fig 8; 123p; English.
CC A cDNA clone (T69545) codes for rat pheromone receptor VNI
CC (W19103), a member of a novel family of presumed 7-transmembrane
CC domain receptors that are evolutionary independent of the odorant
CC receptors of the main olfactory epithelium (MOE). It was isolated
CC by differential screening of cDNA libraries constructed from rat
CC single vomeronasal sensory neurons. A VNI insert in pBluescript
CC has been deposited as ATCC 97294. VNI is expressed only in about
CC 4% of VSN and never in the MOE. cDNA clones for VN2-7 were obtd.
CC by PCR and hybridisation (see also T69546-50, VN2 sequence not
CC given). VN polypeptides (W19103-09) can be expressed in host-
CC vector systems for use in identifying modulators for control of
CC maternal, reproductive and social behavior. to increase fertility,
CC control hormone secretion and to regulate food uptake in humans and
CC animals. A human homologue, Hg25 (T69551), has also been obtd.
SQ Sequence 1386 BP; 357 A; 322 C; 271 G; 435 T;
```

```
Query Match 7.1%; Score 18; DB 1; Length 1386;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 42 TCTGACGCTCAGAACTA 59
DB 1042 TCTGACGCTCAGAACTA 1059
RESULT 2
ID T69545 standard; DNA; 314 BP.
AC T69545;
DE 16-MAR-1994 (first entry)
KW Human brain Expressed Sequence Tag EST01042.
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Cds
FT
FT WO9316178-A.
PD 19-AUG-1993.
PE 12-FEB-1993: U01294.
PR 12-FEB-1992: US-837195.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
```

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:37:13 ; Search time 3268.17 Seconds  
(without alignments)  
212.139 Million cell updates/sec

Title: US-09-049-696-9

Perfect score: 218  
Sequence: 1 GAACAAATGGCCCTCATTGATG.....TCTCTGGGATCCCAATGGAC 218

Scoring table: OLIGO\_MDC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_st:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_v1:\*  
18: em\_fun:\*  
19: em\_htg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_ot:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_v1:\*  
34: gb\_htg1:\*  
35: gb\_htg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pt4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	218	100.0	3311	11	AF039400	Homo sapi

2	216	100.0	2826	42	AF127036	AF127036 Homo sapi
3	148	67.9	35278	11	AF039401	AF039401 Homo sapi
4	26	11.9	2937	12	AB017156	AB017156 Mus muscu
5	21	9.6	191356	42	AC005041	AC005041 Homo sapi
6	19	8.7	169526	35	AC007081	AC007081 Drosophi
7	18	8.3	113143	9	HS544A11	AL023281 Human DNA
8	18	8.3	41505	11	AC004205	AC004205 Homo sapi
9	18	8.3	197915	11	AC005725	AC005725 Homo sapi
10	18	8.3	236822	34	DB4394	DB4394 Homo sapien
11	18	8.3	190000	35	AC006427	AC006427 Homo sapi
12	18	8.3	177294	35	AC007227	AC007227 Homo sapi
13	18	8.3	56479	42	AC006034	AC006034 Homo sapi
14	18	8.3	35491	42	AF048727	AF048727 Homo sapi
15	17	7.8	208430	1	BSIB0005	Z99108 Bacillus su
16	17	7.8	9444	1	D78508	D78508 Bacillus su
17	17	7.8	42861	1	MTCY339	L36547 Pseudomonas
18	17	7.8	24489	1	SPDNAPR	X84072 S.pristine
19	17	7.8	1806	1	SYCSLRB	D64000 Synchocyst
20	17	7.8	146271	1	BOVNEURXIA	L14855 Bos taurus
21	17	7.8	6349	3	FC027351	U27351 Fells catus
22	17	7.8	2534	3	POT3111G	X17255 Bacterioph
23	17	7.8	19680	6	AFAC004411	U63550 Fragaria x
24	17	7.8	106329	8	AB011150	AB011150 Homo sapi
25	17	7.8	1408	9	AB012723	AB012723 Homo sapi
26	17	7.8	5703	9	HS82J11	Z83850 Human DNA s
27	17	7.8	40850	9	HS82J11	AL023581 Human DNA
28	17	7.8	143747	9	HS82J11	AC00399 Human PAC
29	17	7.8	132981	10	HS4514	AC004601 Human 11p
30	17	7.8	137947	11	AC004601	AC004774 Homo sapi
31	17	7.8	133457	11	AC004601	AC005684 Homo sapi
32	17	7.8	132867	11	AC004601	AC005747 Homo sapi
33	17	7.8	161652	11	AC005684	AC005960 Mus muscu
34	17	7.8	143865	11	AC005747	AF001871 Mus muscu
35	17	7.8	16701	11	AC005960	X17502 Mouse mRNA
36	17	7.8	158414	12	AF001871	M66374 Rattus norv
37	17	7.8	1470	12	MMECA39	S76673 mRE3 [mice
38	17	7.8	2453	12	RATNINA	AC003095 Homo sapi
39	17	7.8	5609	12	S76673	AC006712 Caenorhab
40	17	7.8	2469	12	S76673	AL031664 Homo sapi
41	17	7.8	42398	34	AC006712	AC005948 Homo sapi
42	17	7.8	299919	34	HS531H16	AF034627 Phytomona
43	17	7.8	155116	34	AC005948	
44	17	7.8	155093	35	AF034627	
45	17	7.8	210	37	AF034627	

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	MRNA	PRI	15-DEC-1998
LOCUS	AF039400	Homo sapiens	calcium-dependent chloride channel-1 (hCLCA1)	MRNA		
DEFINITION	AF039400	Homo sapiens	calcium-dependent chloride channel-1 (hCLCA1)	MRNA		
ACCESSION	AF039400	94009457	GI:4009457			
VERSION	AF039400.1	GI:4009457				
KEYWORDS		human.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS		Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE		Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Paul,B.U.				
JOURNAL		Genomic cloning, molecular characterization, and functional				
MEDLINE		analysis of human CLCA1, the first human member of the family of				
REFERENCE		Ca2+-activated Cl- channel proteins				
AUTHORS		Genomics 54 (2), 200-214 (1998)				
TITLE		2 (bases 1 to 3311)				
		Gruber,A.D., Elble,R. and Paul,B.U.				
		Direct Submission				

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

## FEATURES

source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/gene="hCLCA1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:g4009458"  
/db\_xref="GI:4009458"  
/translation="MGPKSSVFILILHLEGALSNSLIQLNNGEGIVVAIDPNVP  
EDELIIQIKDMVTOASLYLEATGKRFYFNVAIILPETWTKADYVRPLLEYKNA  
DVLVAESTPPGNDDEPYTEOMNGCEGKERIHLTPFIAGKLAEGPOGKAFFHEMAH  
LRMGVFEYNDDEKYLISNGRIQAVRSAGITGVNKKCOGSGCYTRKCFENKVTGL  
YKCGEFLVRSORPEKASIMRAQHDYSVECTEONHNKRNPNOKNCRSMTEVI  
RDSDFKRTTTPPTPTPTSLQIGRIVCLVLDKSGMATGRRLRLNQAQLEL  
LQVLEGSWGVAVTFDASAHOSELIIQNSGSDRDTAKRLPAASGTSICGLRKA  
FTVLRKYPIDGSEIVLLTDEEDNTISGCFNEVQSGAIHTVALGPSAQLLELSK  
MTGLOQVYASPOVONNGLIDAFGLSSGNGAVSORIOLSKGLTLQNSOMNGTVIY  
DSYKDLFLITWTPOPLLMDPSGOGGVNKRKMAVLOIPIGIAKVGTVK  
SLOASQTLITLTVSRASNATLPITITVSKNKDTSKRPSPPLVYVANNROASPILRA  
SVTLAISVNGKTYTLELDNGAADAADKDGVSRYFTTYDNGRYSVKRALGVN  
AARRVVIPOQSGALYIPGMLENDIQLMPRPPEINKDVQHKQVCFSTSSGGSFVAS  
DVPNAPIDLEPPPGQITDLKAEIHGSLINLTWAPDDYDGHGTAHRYIIRISTSLD  
LRDNFNSLQVNTALIPKEANSEVEFLFKPENTFFENGTDLFIQIADVDRDKSEI  
SNIRVSLFIPQPTPEPSPDETSAQCPNINHISTIPGIHLIKIMMKWIGELQLSIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 218; DB 11; Length 3311;

Best Local Similarity 100.0%; Pred. No. 9,8e-125;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAACAATGGCCTCATTTGATGCTTTGGGCCCCCTTCATCAGGAATGAGCTGTCTCTCA 60  
DB 1743 GAACAATGGCCTCATTTGATGCTTTGGGCCCCCTTCATCAGGAATGAGCTGTCTCTCA 1802  
QY 61 GCGCTCCATCCAGCTTGAAGATTAACCCCTCCAGAACAGCCATGATGATGATG 120  
DB 1803 GCGCTCCATCCAGCTTGAAGATTAACCCCTCCAGAACAGCCATGATGATGATG 1862  
QY 121 CACAGTATCGTGGACAGACCGTGGAAAGACACTTGTTCATACCTGAGCAAC 180  
DB 1863 CACAGTATCGTGGACAGACCGTGGAAAGACACTTGTTCATACCTGAGCAAC 1922  
QY 181 GCAGCCTCCCAATCTCTCTGGGATCCAGTGAC 218  
DB 1923 GCAGCCTCCCAATCTCTCTGGGATCCAGTGAC 1960

## RESULT 2

AF127036 2826 bp mRNA PRI 16-APR-1999

LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL1)  
DEFINITION mRNA, complete cds.

## ACCESSION

AF127036

VERSION 94585468

## KEYWORDS

SOURCE

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
1 (bases 1 to 2826)  
Agnel,M., Verma,T. and Culouscou,J.-M.  
Cloning of three human homologs of bovine epithelial chloride  
channel

JOURNAL Unpublished  
2 (bases 1 to 2826)  
Agnel,M. and Culouscou,J.-M.  
Direct Submission  
Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrieres, Rueil-Malmaison 92500, France

## FEATURES

source  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="CACCL1"  
5. .2749  
/gene="CACCL1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PID:g4585469"  
/db\_xref="GI:4585469"  
/translation="MGPKSSVFILILHLEGALSNSLIQLNNGEGIVVAIDPNVP  
EDELIIQIKDMVTOASLYLEATGKRFYFNVAIILPETWTKADYVRPLLEYKNA  
DVLVAESTPPGNDDEPYTEOMNGCEGKERIHLTPFIAGKLAEGPOGKAFFHEMAH  
LRMGVFEYNDDEKYLISNGRIQAVRSAGITGVNKKCOGSGCYTRKCFENKVTGL  
YKCGEFLVRSORPEKASIMRAQHDYSVECTEONHNKRNPNOKNCRSMTEVI  
RDSDFKRTTTPPTPTPTSLQIGRIVCLVLDKSGMATGRRLRLNQAQLEL  
LQVLEGSWGVAVTFDASAHOSELIIQNSGSDRDTAKRLPAASGTSICGLRKA  
FTVLRKYPIDGSEIVLLTDEEDNTISGCFNEVQSGAIHTVALGPSAQLLELSK  
MTGLOQVYASPOVONNGLIDAFGLSSGNGAVSORIOLSKGLTLQNSOMNGTVIY  
DSYKDLFLITWTPOPLLMDPSGOGGVNKRKMAVLOIPIGIAKVGTVK  
SLOASQTLITLTVSRASNATLPITITVSKNKDTSKRPSPPLVYVANNROASPILRA  
SVTLAISVNGKTYTLELDNGAADAADKDGVSRYFTTYDNGRYSVKRALGVN  
AARRVVIPOQSGALYIPGMLENDIQLMPRPPEINKDVQHKQVCFSTSSGGSFVAS  
DVPNAPIDLEPPPGQITDLKAEIHGSLINLTWAPDDYDGHGTAHRYIIRISTSLD  
LRDNFNSLQVNTALIPKEANSEVEFLFKPENTFFENGTDLFIQIADVDRDKSEI  
SNIRVSLFIPQPTPEPSPDETSAQCPNINHISTIPGIHLIKIMMKWIGELQLSIA  
"

BASE COUNT 875 a 623 c 632 g 656 t  
ORIGIN

Query Match 100.0%; Score 218; DB 42; Length 2826;

Best Local Similarity 100.0%; Pred. No. 9,7e-125;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAACAATGGCCTCATTTGATGCTTTGGGCCCCCTTCATCAGGAATGAGCTGTCTCTCA 60  
DB 1396 GAACAATGGCCTCATTTGATGCTTTGGGCCCCCTTCATCAGGAATGAGCTGTCTCTCA 1455  
QY 61 GCGCTCCATCCAGCTTGAAGATTAACCCCTCCAGAACAGCCATGATGATGATG 120  
DB 1456 GCGCTCCATCCAGCTTGAAGATTAACCCCTCCAGAACAGCCATGATGATGATG 1515  
QY 121 CACAGTATCGTGGACAGACCGTGGAAAGACACTTGTTCATACCTGAGCAAC 180  
DB 1516 CACAGTATCGTGGACAGACCGTGGAAAGACACTTGTTCATACCTGAGCAAC 1575  
QY 181 GCAGCCTCCCAATCTCTCTGGGATCCAGTGAC 218  
DB 1576 GCAGCCTCCCAATCTCTCTGGGATCCAGTGAC 1613

## RESULT 3

AF039401

## LOCUS

## DEFINITION

complete cds.

AF039401 35278 bp DNA PRI 15-DEC-1998

Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.

ACCESSION AF039401  
NTID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Eibler,R.C., Ji,H.L., Schreur,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Eibler,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source 1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="GI:4009460"  
/translation="MGPKSSVFLILHLLEGALNSLIOLNNGYEGIVVAIDPNP  
EDETLIQIKDMYQASLYIEATGKRFYKKNAILLPERWKKADYVRKLTETYNAL  
DVAIVASTIPGNDPEPTEDMGNGGGERHHLPTTAGKKLAEIGPGAKAFVEMAH  
LWGVGDEYNNDKEFLYLSNGRIQAVRCSAGITGTNVYKCGGSCYTKRTFNKVTGL  
YERGEFFVLQSRQTERASIMFAQHDVSIYEFCTEONHNKCAPKQNKCMRSTWEVI  
RSEDEKKTTPMTPTOPNPFSLQIGQIVCVLDKSGSMATNGRNLRLNQGQFL  
LOTVELGSWVGWETPDAAHVOSELIOINSGSDRLAKRLPAASGGTSGSLSA  
FYIRKKYPRDGESEYILTDGEDNTISGCNENKOGATLHTYALPSAOWEJELSK  
MTGGLOTASDOVQNNGLIDAFGALSNGAVQSRSTQLSKSLIQNSOMWEGIVY  
DSTVGRDTFLITWTQPPQIILMDPSGQGGFVVDKNTKMYLIDIPGIAGVGTAK  
SLOASSQTLTLITVSRASNAITLPIITVTSKRTNDFSPPLVYVIANIRGAPILRA  
SVALLIESVNGKVTLEILDNGAGADATKDDGYSRFTTYDINGRVSRYRALGCVN  
AARRVYIPOGSGALYIPGTEENDEIOINPRPINDMDVOHKVCFSSRSSSGSFVAS  
DVPNAPIPDLFPPGOITLDLNAELHGSGLILWTATAGDDYDHTAKYITRISTILD

LDKNESEIOWNTALIPKANESEVLEKRPENITPENGTDLFIAIOAVDKYDLSSEI  
SNARVSLFTPPOTPPETSPDETSAPCPNINHINSTITPGIHILIKMKWIGELQLSIA  
"  
6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 67.9%; Score 148; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 3,9e-81;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 71 CAGCTGAGAGTAAGGATTAACTTCACAGACGCCAGTGATGATGACAGTGATC 130  
|||||  
Db 26641 CAGCTGAGAGTAAGGATTAACTTCACAGACGCCAGTGATGATGACAGTGATC 26700  
QY 131 GTGACAGCACCGTGGGAAGGACACTTGTCTTATACCTCGGACCAAGCAGCCGCC 190  
|||||  
Db 26701 GTGACAGCACCGTGGGAAGGACACTTGTCTTATACCTCGGACCAAGCAGCCGCC 26760  
QY 191 CAATCCTCTCTGGGATCCCACTGAC 218  
|||||  
Db 26761 CAATCCTCTCTGGGATCCCACTGAC 26788  
RESULT 4  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS AB017156  
ACCESSION AB017156  
NID 93721911  
VERSION AB017156.1 GI:3721911  
KEYWORDS GOB-5.  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 191356)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciironathi; Muridae; Murinae; Mus.  
 TITLE  
 AUTHOR  
 JOURNAL  
 Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice  
 TITLE  
 AUTHOR  
 JOURNAL  
 Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
 99160866  
 2 (bases 1 to 2937)  
 Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
 Direct Submission  
 Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru Komiyama, ERARO, Jst, Hirohashi Cell Configuration Project; 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan  
 (E-mail:tkom@ccp.jst.go.jp, Tel:81-298-47-7563, Fax:81-298-47-5226)  
 FEATURES  
 source  
 Location/Qualifiers  
 1..2937  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_type="goblet cell"  
 /dev\_stage="adult"  
 /tissue\_type="intestine"  
 15..2756  
 /gene="gob-5"  
 15..2756  
 /gene="gob-5"  
 /codon\_start=1  
 /protein\_id="BAA33743.1"  
 /db\_xref="PID:d1034712"  
 /db\_xref="PID:g3721912"  
 /db\_xref="GI:3721912"  
 /translation="MESLSPYELLILHLLEGVLSLQLNNNGEGIVLIDHVP  
 EPEALIOHKDWTOASPYLFATGKRFYKFNVALIPESMKAPEYTPKLETERNA  
 DYLVSTSPLENDERTHEHICAGEKGIHILPDLAKKILNOYGPDRFVHEMAH  
 FRMGVNEYNDKEKFLSKGRPOAVKCAITTKGNVRCOGSCCTTKCYIDRYTG  
 LYKDNCEVFPDHPQNEKASIMFNQINSVECTEKNNQENLNDONQCNLSRWEY  
 IOESDEKQITTPATQAPAPATSLQIGRIVCLVYDKSGSMLNDRLNRNMQASLFL  
 LIOTVEGSGWGVAFDPSAAYVOSELKQNSGADRLILKILPTVSAGTSLCSGLRT  
 ATVLKIKKYPGSEIYLJLDGDNITSCFEDLVKQSGAIHIVLALGPAAKELELS  
 KMGIGQITSSPOVONNGNVDAPFALISGNAIAHISIOLESRGVNLONNMNGSVI  
 VSSVGGKDTLFLTTHTPTIFIMPDSVEONGEFLDPTTAYAYOVGCTAKVPEPK  
 YSLQASSQTLTLTVTSRASATLPTIVFPVKNKNGKPPSPYTYASIRQASPIIR  
 ASVTALIESVNGKTVLELIDNGAGADAKNGVYSRFTAFDANRISVKIWLGGV  
 TSDRORAPPKNRAMIDQWIEDGEVRAMPPREISYVODKQCSRTSGSGSVATN  
 VPAAPIDPLEPCOITDLKASIOGONLVNLWTAGDDYDGRASNYIIRMSISIVD  
 LDHNTSLQVNTGLIPKASSEIIFEEELGNTGNGDIFIAIOAVDKSNLKEI  
 SNIRASVTPRADEPRIPEDSPPCDISINSTIPGIIHVKIMMKVLGMYOTLIGH"

REFERENCE 1 (bases 1 to 191356)  
 Waterston, R.H.  
 The sequence of Homo sapiens clone  
 unpublished  
 2 (bases 1 to 191356)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 3 (bases 1 to 191356)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (24-MAR-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Mar 24, 1999 this sequence version replaced g1:3309088.  
 COMMENT  
 FEATURES  
 source  
 Location/Qualifiers  
 1..191356  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NH0523H20"  
 BASE COUNT  
 46280 a 47882 c 48007 g 49187 t  
 ORIGIN  
 Query Match 9.6%; Score 21; DB 42; Length 191356;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 182 CAGCTCCCAATCTCTC 202  
 DB 4335 CAGCTCCCAATCTCTC 4335  
 RESULT 6  
 AC007081/c HTG 07-JUN-1999  
 LOCUS  
 DEFINITION  
 Drosophila melanogaster chromosome 3 clone BACR48023 (D98) RPOI-98  
 48,0.23 map 99A7-99B5 strain y; cn bw sp, WORKING DRAFT SEQUENCE, 3  
 uncoded pieces.  
 AC007081  
 94934937  
 NID  
 VERSION  
 AC007081.5 GI:4934937  
 KEYWORDS  
 HTG; HTGS\_PHASE1.  
 SOURCE  
 ORGANISM  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 169526)  
 Celniker, S.E., Agbayan, A., Arcaluna, T.T., Baxter, E., Blazet, R.G.,  
 Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
 Hoskins, R.A., Houston, K.A., Hummstli, S.B., Karra, K., Kearney, L.,  
 Kim, E., Lee, B., Lewis, S., Li, P., Lomtan, M.A., Mazda, P.,  
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclob, J.M., Park, S.,  
 Pfeiffer, B., Poon, L., Segueira, A., Sethi, H., Shif, E.,  
 Svitskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
 Rubin, G.M.  
 Sequencing of Drosophila melanogaster  
 Unpublished  
 2 (bases 1 to 169526)  
 Celniker, S.E., Agbayan, A., Arcaluna, T.T., Baxter, E., Blazet, R.G.,  
 Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
 Hoskins, R.A., Houston, K.A., Hummstli, S.B., Karra, K., Kearney, L.,  
 Kim, E., Lee, B., Lewis, S., Li, P., Lomtan, M.A., Mazda, P.,  
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclob, J.M., Park, S.,  
 Pfeiffer, B., Poon, L., Segueira, A., Sethi, H., Shif, E.,  
 Svitskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
 Rubin, G.M.  
 Direct Submission

## JOURNAL

## COMMENT

Submitted (16-MAR-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On Jun 1, 1999 this sequence version replaced gi:4803907.  
 For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [dbgap@fruitfly.berkeley.edu](mailto:dbgap@fruitfly.berkeley.edu). All contigs in this submission meet the following cutoffs: length >= 400 bases, phrap computed error rate <= 1/10.

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence \* as soon as it is available and the accession number will be preserved.

## FEATURES

## SOURCE

1 733: contig of 733 bp in length  
 \* 734 813: gap of unknown length  
 \* 814 41210: contig of 40397 bp in length  
 \* 41211 41290: gap of unknown length  
 \* 41291 169526: contig of 128236 bp in length.  
 Location/Qualifiers

1.169526  
 /organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/clone="BACR48023 (D498) RPCI-98 48.0.23"

/chromosome="3"

/clone\_1lb="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcorI in

BAC3.6"

/map="99A7-99B5"

BASE COUNT 47509 a 37630 c 37526 g 46700 t 161 others

## ORIGIN

## Query Match

Best Local Similarity 8.7%; Score 19; DB 35; Length 169526;  
 Best Local Similarity 100.0%; Pred. No. 0.75;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TTGATGCTTTGGGCCCT 33

Db 81329 TTGATGCTTTGGGCCCT 81311

## RESULT 7

## LOCUS

HS544A11 113143 bp DNA PRI 28-JAN-1999

## DEFINITION

Human DNA sequence from clone 544A11 on chromosome 22q12.1-12.3.

## ACCESSION

## NID

## KEYWORDS

## SOURCE

## ORGANISM

## TITLE

## JOURNAL

## COMMENT

## AUTHORS

## JOURNAL

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## FEATURES

## SOURCE

once, or longer because we arrange for a small overlap between neighbouring submissions.  
 The true left end of clone B42B1 (Z75890) is at 21055 in this sequence. The true left end of clone 732B4 (AL008722) is at 113044. The true right end of clone B42B1 is at 62012.  
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
 This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>  
 544A11 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218. VECTOR: pBelBAC11.

Location/Qualifiers

1.113143  
 /organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="22"

/clone="544A11"

/map="q12.1-12.3"

1.4047  
 /note="U1 repeat: matches 1341. .5390 of consensus"

3900. .4787  
 /note="L1PAP2 repeat: matches 3. .891 of consensus"

5114. .5173  
 /note="10 copies 6 mer ggtgtt 80% conserved"

5606. .5678  
 /note="MIR2 repeat: matches 68. .146 of consensus"

6710. .6783  
 /note="AluYo repeat: matches 289. .1 of consensus"

6710. .6783  
 /note="L1MA2 repeat: matches 976. .1055 of consensus"

8614. .8732  
 /note="MER5A repeat: matches 28. .1 of consensus"

8733. .9031  
 /note="MIR repeat: matches 49. .167 of consensus"

8733. .9031  
 /note="AluX repeat: matches 3. .302 of consensus"

11345. .11482  
 /note="MIR2 repeat: matches 1. .137 of consensus"

13487. .13705  
 /note="MER20 repeat: matches 1. .217 of consensus"

14597. .14880  
 /note="MER5A repeat: matches 101. .26 of consensus"

14881. .14928  
 /note="AluSg repeat: matches 1. .300 of consensus"

14881. .14928  
 /note="8 copies 6 mer gaagaa 81% conserved"

15914  
 /note="L1M1 repeat: matches 914. .133 of consensus"

15929. .16229  
 /note="AluX repeat: matches 301. .2 of consensus"

16700  
 /note="AluYo repeat: matches 302. .37 of consensus"

16705. .17109  
 /note="Alu repeat: matches 4941. .5371 of consensus"

17033. .17903  
 /note="L1 repeat: matches 53. .931 of consensus"

17923. .18223  
 /note="L1MB3 repeat: matches 53. .931 of consensus"

18379. .18679  
 /note="AluX repeat: matches 1. .302 of consensus"

19179. .19715  
 /note="AluSg repeat: matches 1. .300 of consensus"

20860. .20899  
 /note="match: GSS B14146"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

22323. .22619  
 /note="20 copies 2 mer tt 83% conserved"

```

repeat_region /note="Alusg repeat: matches 1. .296 of consensus"
                /note="Alusg repeat: matches 294. .130 of consensus"
repeat_region /note="Alusg repeat: matches 298. .1 of consensus"
                /note="Alusg repeat: matches 133. .9 of consensus"
repeat_region /note="Alusg repeat: matches 299. .1 of consensus"
                /note="Alusg repeat: matches 299. .1 of consensus"
variation      /note="clone CB42B1; att in this entry; substitution"
                /replace="act"
repeat_region /note="MIR repeat: matches 144. .62 of consensus"
                /note="MIR repeat: matches 144. .62 of consensus"
variation      /note="clone CB42B1; ta in this entry; deletion"
                /replace="taa"
repeat_region /note="MIR repeat: matches 4892. .5022 of consensus"
                /note="MIR repeat: matches 4892. .5022 of consensus"
variation      /note="clone CB42B1; ata in this entry; substitution"
                /replace="aaa"
repeat_region /note="MLT2_internal repeat: matches 290. .431 of
                /note="MLT2_internal repeat: matches 290. .431 of
                consensus"
repeat_region /note="L1 repeat: matches 1981. .1700 of consensus"
                /note="L1 repeat: matches 1981. .1700 of consensus"
repeat_region /note="MER3 repeat: matches 1. .113 of consensus"
                /note="MER3 repeat: matches 1. .113 of consensus"
repeat_region /note="MLT2_internal repeat: matches 547. .625 of
                /note="MLT2_internal repeat: matches 547. .625 of
                consensus"
repeat_region /note="MER20 repeat: matches 218. .98 of consensus"
                /note="MER20 repeat: matches 218. .98 of consensus"
repeat_region /note="L1MC3 repeat: matches 987. .1354 of consensus"
                /note="L1MC3 repeat: matches 987. .1354 of consensus"
repeat_region /note="MER42c repeat: matches 2. .168 of consensus"
                /note="MER42c repeat: matches 2. .168 of consensus"
repeat_region /note="MER42B repeat: matches 166. .1204 of consensus"
                /note="MER42B repeat: matches 166. .1204 of consensus"
repeat_region /note="MER42c repeat: matches 758. .1538 of consensus"
                /note="MER42c repeat: matches 758. .1538 of consensus"
repeat_region /note="Alusg repeat: matches 1. .301 of consensus"
                /note="Alusg repeat: matches 1. .301 of consensus"
variation      /note="clone CB42B1; aaaaaaaaaa in this entry;
                /note="clone CB42B1; aaaaaaaaaa in this entry;
                insertion"
                /replace="aa"
repeat_region /note="MIR2 repeat: matches 145. .1 of consensus"
                /note="MIR2 repeat: matches 145. .1 of consensus"
prim_transcript 33749. .34016
                /note="match: EST AA313175"
repeat_region /note="MIR repeat: matches 203. .120 of consensus"
                /note="MIR repeat: matches 203. .120 of consensus"
repeat_region /note="MIR repeat: matches 127. .256 of consensus"
                /note="MIR repeat: matches 127. .256 of consensus"
repeat_region /note="MIR repeat: matches 210. .28 of consensus"
                /note="MIR repeat: matches 210. .28 of consensus"
repeat_region /note="MIR repeat: matches 1. .262 of consensus"
                /note="MIR repeat: matches 1. .262 of consensus"
repeat_region /note="MER2 repeat: matches 13. .344 of consensus"
                /note="MER2 repeat: matches 13. .344 of consensus"
variation      /note="clone CB42B1; ata in this entry; substitution"
                /replace="aca"
repeat_region /note="MIR repeat: matches 84. .119 of consensus"
                /note="MIR repeat: matches 84. .119 of consensus"
repeat_region /note="MIR repeat: matches 1. .141 of consensus"
                /note="MIR repeat: matches 1. .141 of consensus"
repeat_region /note="Alusg repeat: matches 1. .302 of consensus"
                /note="Alusg repeat: matches 1. .302 of consensus"
repeat_region /note="Alusg repeat: matches 3. .298 of consensus"
                /note="Alusg repeat: matches 3. .298 of consensus"

```

```

variation      45200. .45201
                /note="clone CB42B1; ag in this entry; deletion"
                /replace="aag"
prim_transcript 45385. .45626
                /note="match: EST T91723"
repeat_region 45648. .45943
                /note="Alusg repeat: matches 5. .301 of consensus"
repeat_region 46014. .46225
                /note="L1 repeat: matches 4565. .4778 of consensus"
repeat_region 46262. .46561
                /note="L1 repeat: matches 4565. .4778 of consensus"
repeat_region 46573. .47151
                /note="Alusg repeat: matches 1. .300 of consensus"
repeat_region 46573. .47151
                /note="L1MC4 repeat: matches 208. .789 of consensus"
repeat_region 47150. .47218
                /note="L1MC3 repeat: matches 834. .898 of consensus"
repeat_region 47242. .47285
                /note="L2 copies 2 mer at 98% conserved"
repeat_region 47244. .47285
                /note="7 copies 6 mer atatat 100% conserved"
variation      47280. .47285
                /note="clone CB42B1; atatat in this entry; insertion"
                /replace="at"
variation      47414. .47416
                /note="clone CB42B1; gtt in this entry; substitution"
                /replace="gct"
repeat_region 48149. .48393
                /note="L1MC2 repeat: matches 77. .334 of consensus"
repeat_region 48407. .48436
                /note="L1 copies 2 mer gt 100% conserved"

Query Match      8.3%; Score 18; DB 9; Length 13143;
Best Local Similarity 100.0%; Pred. NO. 3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 CAGTGGATGATGGCACA 124
DB 75407 CAGTGGATGATGGCACA 75424

RESULT 8
AC004205 41506 bp DNA PRI 25-NOV-1998
LOCUS Homo sapiens clone UMC:Y3C071 from 6p21, complete sequence.
DEFINITION AC004205
ACCESSION 93927854
NID AC004205.1 GI:3927854
VERSION 1
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens.
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 41506)
AUTHORS Gullandux,T., Janer,M., Wong,G.K.Y., Spies,T. and Geraghty,D.E.
TITLE The complete genomic sequence of 424,015 bp at the centromeric end
of the H1A class I region: gene content and polymorphism
Proc. Natl. Acad. Sci. U.S.A. 95 (16), 9494-9499 (1998)
JOURNAL 98356185
MEDLINE Fred Hutchinson Cancer Research Center
REMARK The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
Contact: Daniel E. Geraghty (geraghty@fhrc.org)
2 (bases 1 to 41506)
REFERENCE Geraghty,D.E. and Olson,M.V.
AUTHORS Direct Submission
TITLE Submitted (23-FEB-1998) Human Genome Center, University of
JOURNAL Washington, Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 41506)
REFERENCE Geraghty,D.E. and Olson,M.V.
AUTHORS Direct Submission
TITLE Submitted (25-NOV-1998) Human Genome Center, University of
JOURNAL Washington, Box 352145, Seattle, WA 98195, USA
COMMENT On Nov 25, 1998 this sequence version replaced gi:2905855.

```



## Overlapping Sequences:

5' : UMGc:y3c028 (Genbank Accession: AC004210)  
3' : UMGc:y3c018 (Genbank Accession: AC004204)

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 76.68  
DS or two chemistry coverage: 99.98  
Single stranded regions: 2

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

Map	Seq	Map	Seq	Map	Seq
BglII		HindIII		NsiI	
6493.25	6534.00	2906.67	2931.00	965.88	958.00
10020.00	9916.00	2230.67	2225.00	4984.38	4948.00
756.25	755.00	3744.62	3714.00	675.00	674.00
1474.00	1475.00	1451.75	1453.00	3790.75	3762.00
2503.75	2519.00	1023.25	1015.00	3344.25	3327.00
5066.00	5111.00	2795.86	2795.00	1157.75	1152.00
2344.00	2361.00	2232.50	2258.00	2747.00	2747.00
1916.00	1938.00	5171.00	5176.00	677.67	682.00
5301.00	5219.00	1910.00	1924.00	677.67	667.00
981.00	972.00	1441.67	1447.00	4056.00	4038.00
1253.00	1266.00	1093.00	1095.00		
		9094.50	8842.00		
		841.00	831.00		

## FEATURES

## source

Location/Qualifiers  
1. 41506  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21"  
/clone="CGM1:B38D3"  
/sub\_clone="UMGC:y3c071"  
/cell\_line="CGM1"  
/clone\_lib="Wash U YAC library"  
10. 227  
/rpt\_family="Alu"  
repeat\_region  
1270. 1504  
/rpt\_family="MER25"  
repeat\_region  
complement(1837..2123)  
/rpt\_family="Alu"  
2135. 2300  
repeat\_region

repeat\_region /rpt\_family="LI"  
complement(2581..3065)  
repeat\_region /rpt\_family="MER9"  
3281..4199  
repeat\_region /rpt\_family="LI"  
complement(4733..4871)  
repeat\_region /rpt\_family="LRR10"  
11921..12186  
repeat\_region /rpt\_family="Alu"  
complement(13151..13371)  
repeat\_region /rpt\_family="LRR10"  
13442..13560  
repeat\_region /rpt\_family="LI"  
complement(13919..19370)  
repeat\_region /rpt\_family="LI"  
15112..15394  
repeat\_region /rpt\_family="Alu"  
15524..15820  
repeat\_region /rpt\_family="Alu"  
17260..17551  
repeat\_region /rpt\_family="Alu"  
complement(19213..19447)  
repeat\_region /rpt\_family="MER25"  
complement(19538..19648)  
repeat\_region /rpt\_family="LI"  
20344..20620  
repeat\_region /rpt\_family="Alu"  
21823..22104  
repeat\_region /rpt\_family="Alu"  
23830..23893  
repeat\_region /rpt\_family="MLT1"  
25591..25687  
repeat\_region /rpt\_family="Alu"  
complement(26880..27051)  
repeat\_region /rpt\_family="MER3"  
30616..30864  
repeat\_region /rpt\_family="Alu"  
complement(34535..34834)  
repeat\_region /rpt\_family="Alu"  
complement(35723..35832)  
repeat\_region /rpt\_family="MIR"  
38630..38916  
repeat\_region /rpt\_family="Alu"  
38943..39423  
repeat\_region /rpt\_family="LI"  
BASE COUNT 12317 a 8781 c 8500 g 11908 t  
ORIGIN

Query Match 8.3%; Score 18; DB 11; Length 41506;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 195 TCCTCTCTGCGATCCCA 212  
Db 40014 TCCTCTCTGCGATCCCA 40031

RESULT 9  
LOCUS AC005725 197915 bp DNA PRI 02-DEC-1998  
DEFINITION Homo sapiens chromosome 17, clone hRPK.680.C\_21, complete sequence.  
ACCESSION AC005725  
NID 93947428  
VERSION AC005725.1 GI:3947428  
KEYWORDS HTG  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 197915)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 17, clone hRPK.680.C\_21

```

JOURNAL REFERENCE
AUTHORS Unpublished
2 (bases 1 to 197915)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraldery,K., Grant,G., Hagos,B., Hearford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Margulis,N., McEwan,P.,
McGuirk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J.,
Mychalckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Testaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.

TITLE
JOURNAL Direct Submission
REFERENCE Submitted (29-SEP-1998) Whitehead Institute/MIT Center for Genome
AUTHORS Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 197915)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraldery,K., Grant,G., Hagos,B., Hearford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Margulis,N., McEwan,P.,
McGuirk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J.,
Mychalckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Testaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.

TITLE
JOURNAL Direct Submission
REFERENCE Submitted (02-DEC-1998) Whitehead Institute/MIT Center for Genome
AUTHORS Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 2, 1998 this sequence version replaced qt:3935222.
All repeats were identified using RepeatMasker: Smtt, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT
Only the last 197915 base pairs of this clone are being submitted.
The remainder overlaps accession number AC005548 (MIGR project
L361).

FEATURES
source
1. 197915
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK.680.C.21"
/clone_lib="RPCT-11 human BAC library"
/map="17"
/chromosome="17"
complement(251..355)
/rpt_family="MIR"
repeat_region
/rpt_family="AluSc"
687..969
/rpt_family="AT_rich"
repeat_region
/rpt_family="AT_rich"
2640..2661
/rpt_family="AluSc"
4789..5192
/rpt_family="MLT1B"
repeat_region
complement(5330..5795)
/rpt_family="MLT1H"
repeat_region
complement(7650..7676)
/rpt_family="(TGCA)n"
repeat_region
complement(7722..8030)
/rpt_family="L1R16A"
repeat_region
complement(8051..8475)
/rpt_family="L2"
repeat_region
complement(8580..8676)
/rpt_family="MIR"

```

```

repeat_region complement(8866..8935)
/rpt_family="MLT1H"
repeat_region 8936..8995
/rpt_family="(TAGA)n"
repeat_region complement(8998..9054)
/rpt_family="(TGCA)n"
repeat_region complement(9057..9381)
/rpt_family="MLT1H"
repeat_region 9391..9691
/rpt_family="AluSc"
repeat_region 9756..9824
/rpt_family="Polypurine"
complement(9825..9847)
/rpt_family="MLT1J"
repeat_region 9848..9966
/rpt_family="(GAA)n"
repeat_region complement(9967..10060)
/rpt_family="MLT1J"
repeat_region 10947..11128
/rpt_family="L2"
repeat_region 11187..11570
/rpt_family="L2"
repeat_region complement(12968..13015)
/rpt_family="(CGG)n"
repeat_region 13110..13159
/rpt_family="GC_rich"
repeat_region complement(13396..13420)
/rpt_family="(CGG)n"
repeat_region complement(13586..13654)
/rpt_family="(CGG)n"
repeat_region 13667..13758
/rpt_family="GC_rich"
repeat_region 13864..13927
/rpt_family="GC_rich"
repeat_region 13973..14023
/rpt_family="GC_rich"
repeat_region 18546..18848
/rpt_family="AluY"
repeat_region complement(20108..20409)
/rpt_family="AluY"
repeat_region complement(20687..20805)
/rpt_family="MIR"
repeat_region 21194..21495
/rpt_family="AluY"
repeat_region complement(21620..21682)
/rpt_family="(CA)n"
repeat_region complement(23260..23416)
/rpt_family="MIR"
repeat_region 26556..26614
/rpt_family="MIR"
repeat_region 27378..27513
/rpt_family="L2"
repeat_region 27644..27727
/rpt_family="purine-rich"
repeat_region complement(28209..28257)
/rpt_family="L2"
complement(28225..28485)
/rpt_family="MIR"
repeat_region 28549..28847
/rpt_family="AluSc"
repeat_region 29715..30026
/rpt_family="AluSc"
repeat_region 30027..30068
/rpt_family="polypurine"
repeat_region 30328..30353
/rpt_family="AT_rich"
repeat_region 31266..31388
/rpt_family="(TGCA)n"
repeat_region 31474..31520
/rpt_family="purine-rich"
repeat_region 31537..31784
/rpt_family="MIR"
repeat_region complement(33067..33188)

```

```

repeat_region /rpt_family="purine-rich"
complement(33831..34379)
/rpt_family="LMB6"
repeat_region complement(34377..34527)
/rpt_family="L1"
repeat_region complement(36437..36503)
/rpt_family="MIR"
repeat_region complement(36546..36745)
/rpt_family="MIR"
repeat_region 37422..37595
/rpt_family="MERSB"
complement(37647..37836)
/rpt_family="MERSB"
repeat_region 38648..38746
/rpt_family="MERSA"
repeat_region 38818..39002
/rpt_family="MIR"
complement(39298..39373)
/rpt_family="MIR"
repeat_region 39598..39749
/rpt_family="MERSA"
complement(39764..39885)
/rpt_family="MERSB"
repeat_region 40689..41149
/rpt_family="L2"
repeat_region complement(41364..41675)
/rpt_family="AluUb"
repeat_region 42215..42272
/rpt_family="MER3"
complement(42297..42607)
/rpt_family="AluSg"
repeat_region complement(42851..43091)
/rpt_family="MER58A"
repeat_region 43312..43376
/rpt_family="MLT2D"
complement(43767..44095)
/rpt_family="AluSg"
repeat_region 44800..44932
/rpt_family="MIR"
complement(45393..45589)
/rpt_family="MIR"
repeat_region

```

Query Match 8.3% Score 18; DB 11; Length 197915;  
 Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 195 TCCTCTCTGGATCCCA 212  
 DB 43157 TCCTCTCTGGATCCCA 43174

```

RESULT 10
LOCUS D84394 236822 bp DNA HTG 16-SEP-1997
DEFINITION Homo sapiens chromosome 6 clone y109 map 6p21.3, WORKING DRAFT
ACCESSION D84394
NID 92401262
VERSION D84394.1 GI:2401262
KEYWORDS HTG; HTGS_PHASE2
SOURCE Homo sapiens cell_line:CGM1 DNA, clone_11b:T.Imai and M.V.Olson
clone:y109.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
AUTHORS Inoko,H.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1996) to the DDBJ/EMBL/GenBank databases.
Hidetoshi Inoko, Tokai University School of Medicine, Molecular
Life Science; Bohseida, Isehara, Kanagawa 259-11, Japan
(E-mail:hinoko@is.tcc.u-tokai.ac.jp, Tel:0463-93-1121,
Fax:0463-94-8884)

```

```

REFERENCE 2 (bases 1 to 236822)
AUTHORS Mizuki,N., Ando,H., Kimura,M., Ohno,S., Miyata,S., Yamazaki,M.,
Tashiro,H., Watanabe,K., Ono,A., Taguchi,S., Sugawara,C.,
Fukuzumi,Y., Sato,M., Okumura,K., Goto,K., Ishihara,M., Kikuchi,Y.,
Shina,T., Ando,A., Ikemura,T., and Inoko,H.
TITLE Nucleotide sequence analysis of the HLA class I region spanning the
237 kb segment around the HLA-B and -C genes
JOURNAL Unpublished (1996)
AUTHORS 3 (sites)
Mizuki,N., Ando,H., Kimura,M., Ohno,S., Miyata,S., Yamazaki,M.,
Tashiro,H., Watanabe,K., Ono,A., Taguchi,S., Sugawara,C.,
Fukuzumi,Y., Okumura,K., Goto,K., Ishihara,M., Nakamura,S.,
Yonemoto,J., Kikuchi,Y., Shina,T., Chen,L., Ando,A., Ikemura,T.,
and Inoko,H.

```

```

TITLE Nucleotide sequence analysis of the HLA class I region spanning the
237-kb segment around the HLA-B and -C genes
JOURNAL Genomics 42 (1), 55-66 (1997)
MEDLINE 97321045
COMMENT * NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

#### FEATURES

```

source 1..236822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="y109"
/clone_11b="T.Imai and M.V.Olson"
/map="6p21.3"
BASE COUNT 66997 a 52352 c 52463 g 65010 t
ORIGIN

```

Query Match 8.3% Score 18; DB 34; Length 236822;  
 Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 195 TCCTCTCTGGATCCCA 212  
 DB 112938 TCCTCTCTGGATCCCA 112955

```

RESULT 11
LOCUS AC006427/c 190000 bp DNA HTG 08-JUN-1999
DEFINITION Homo sapiens chromosome 4, WORKING DRAFT SEQUENCE, 12 unordered
pieces.
ACCESSION AC006427
NID 95019269
VERSION AC006427.4 GI:5019269
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission
JOURNAL Unpublished
JOURNAL Submitted (21-JAN-1999) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
On Jun 8, 1999 this sequence version replaced gi:4966377.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

\* as soon as it is available and the accession number will  
\* be preserved.  
1 1122: contig of 1122 bp in length  
\* 1123 2793: gap of unknown length  
\* 2794 3969: contig of 1176 bp in length  
\* 3970 5639: gap of unknown length  
\* 5640 7494: contig of 1855 bp in length  
\* 7495 9164: gap of unknown length  
\* 9165 11251: contig of 2087 bp in length  
\* 11252 12921: gap of unknown length  
\* 12922 24188: contig of 11267 bp in length  
\* 24189 25858: gap of unknown length  
\* 25859 33031: contig of 7173 bp in length  
\* 33032 34701: gap of unknown length  
\* 34702 46041: contig of 11340 bp in length  
\* 46042 47711: gap of unknown length  
\* 47712 67667: contig of 19956 bp in length  
\* 67668 69337: gap of unknown length  
\* 69338 89177: contig of 19840 bp in length  
\* 89178 90847: gap of unknown length  
\* 90848 118409: contig of 27562 bp in length  
\* 118410 120079: gap of unknown length  
\* 120080 148176: contig of 28097 bp in length  
\* 148177 149846: gap of unknown length  
\* 149847 190000: contig of 40154 bp in length.  
Location/Qualifiers  
1..190000  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
BASE COUNT 52922 a 35299 c 34931 g 48469 t 18379 others  
ORIGIN  
Query Match 8.3%: Score 18; DB 35; Length 190000;  
Best Local Similarity 100.0%: Pred. No. 3.2; 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 35 TCATCAGAAATGAGACT 52  
|||||  
Db 178254 TCATCAGAAATGAGACT 178237  
RESULT 12  
AC007227/c 177294 bp DNA HTG 06-APR-1999  
LOCUS Homo sapiens chromosome 16 clone 54H19, WORKING DRAFT SEQUENCE, 17  
DEFINITION  
AC007227  
AC007227  
AC007227.1 GI:4567161  
NID 94567161  
VERSION AC007227.1  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 177294)  
Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,  
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,  
Bryant, J., Tesmer, J., Melnick, L., Longmire, J., White, S., Tatum, O.,  
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,  
Han, C. and Deaven, L.  
Sequencing of Human Chromosome 16p13.3  
Unpublished  
2 (bases 1 to 177294)  
Ricke, D.O.  
Large Scale Sequence Analysis and Annotation with the Sequence  
Comparison Analysis (SCAN) System  
Unpublished  
3 (bases 1 to 177294)  
Jones, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,  
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,  
Bryant, J., Tesmer, J., Melnick, L., Longmire, J., White, S., Tatum, O.,  
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,  
Han, C. and Deaven, L.

Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,  
McMurry, R., Han, C. and Deaven, L.  
Direct Submission  
Submitted (06-APR-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los  
Alamos, NM 87545, USA  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 690: contig of 690 bp in length  
\* 691 1381: gap of unknown length  
\* 1382 2511: contig of 691 bp in length  
\* 2512 3457: gap of unknown length  
\* 3458 4261: contig of 1130 bp in length  
\* 4262 5455: gap of unknown length  
\* 5456 6492: contig of 946 bp in length  
\* 6493 6917: gap of unknown length  
\* 6918 7518: contig of 425 bp in length  
\* 7519 7641: gap of unknown length  
\* 7642 8312: contig of 601 bp in length  
\* 8313 9012: gap of unknown length  
\* 9013 9661: contig of 123 bp in length  
\* 9662 12503: gap of unknown length  
\* 12504 49377: gap of 2842 bp in length  
\* 49378 97428: gap of unknown length  
\* 97429 177294: contig of 36874 bp in length  
\* 177295: gap of 46051 bp in length  
\* 177296: gap of unknown length  
\* 177297: gap of 79866 bp in length.  
Location/Qualifiers  
1..177294  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="54H19"  
/chromosome="16"  
BASE COUNT 42026 a 44693 c 44677 g 45876 t 22 others  
ORIGIN  
Query Match 8.3%: Score 18; DB 35; Length 177294;  
Best Local Similarity 100.0%: Pred. No. 3.1; 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 194 ATCCTTCCTGAGATCC 211  
|||||  
Db 164209 ATCCTTCCTGAGATCC 164192  
RESULT 13  
AC006034/c 56479 bp DNA PRI 14-MAY-1999  
LOCUS Homo sapiens clone NH0144M13, complete sequence.  
AC006034  
AC006034  
AC006034  
NID 94827318  
VERSION AC006034.2 GI:4827318

KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata; Mammalia;  
Eutheria: Primates: Catarrhini; Homnidae; Homo.  
1 (bases 1 to 56479)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 56479)  
TITLE Waterston,R.H.  
REFERENCE Direct Submission  
JOURNAL Submitted (22-NOV-1998) Genome Sequencing Center, Washington  
MO 63108, USA  
3 (bases 1 to 56479)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (14-MAY-1999) Genome Sequencing Center, Washington  
MO 63108, USA  
COMMENT On May 14, 1999 this sequence version replaced gt:419982.  
FEATURES  
SOURCE location/Qualifiers  
1..56479  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NH0144M13"  
BASE COUNT 18089 a 9723 c 10383 g 18284 t  
ORIGIN

Query Match 8.3% Score 18; DB 42; Length 56479;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CTTTCATCAGGAATGCA 49  
|||||  
Db 15673 CTTTCATCAGGAATGCA 15656

RESULT 14  
LOCUS AF048727/c 35491 bp DNA PRI 11-MAR-1999  
DEFINITION Homo sapiens minisatellite cebi repeat region.  
ACCESSION AF048727  
NID 92935483  
VERSION AF048727.1 GI:2935483  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota: Metazoa: Chordata: Craniata: Vertebrata; Mammalia;  
Eutheria: Primates: Catarrhini; Homnidae; Homo.  
1 (bases 1 to 35491)  
AUTHORS Murray,T., Buard,J., Neil,D.L., Yeraman,E., Tamaki,K., Hollis,C.  
and Jeffreys,A.J.  
TITLE Comparative sequence analysis of human minisatellites showing  
meiotic repeat instability  
JOURNAL Genome Res. 9 (2), 130-136 (1999)  
MEDLINE 99148136  
REFERENCE 2 (bases 1 to 35491)  
AUTHORS Buard,J.  
TITLE Sequence from cosmid containing human hypervariable minisatellite  
cebi  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 35491)  
AUTHORS Buard,J.  
TITLE Direct Submission  
JOURNAL Submitted (17-FEB-1998) Genetics, University of Leicester,  
University Road, Leicester LE1 7RH, England  
location/Qualifiers  
1..35491  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/chromosome="2"  
/map="2q37.3"  
/note="from cosmid 53: G. Vergnaud, et al, Genomics,  
11:135-144, 1991"  
13155..13391  
repeat\_region  
/note="minisatellite cebi; repeat unit varies in length  
from 39 to 43 nt  
/rpt\_unit="13156..13204"  
BASE COUNT 8576 a 10002 c 9607 g 7253 t 53 others  
ORIGIN

Query Match 8.3% Score 18; DB 42; Length 35491;  
Best Local Similarity 100.0%; Pred. No. 2.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 AGGCACTTGTTCTTA 167  
|||||  
Db 18430 AGGCACTTGTTCTTA 18413

RESULT 15  
LOCUS BSUB0005/c 208430 bp DNA BCT 26-NOV-1997  
DEFINITION Bacillus subtilis complete genome (section 5 of 21): from 802821 to  
1011250.  
ACCESSION 299108 AL009126  
NID 92633055  
VERSION 299108.1 GI:2633055  
KEYWORDS  
SOURCE Bacillus subtilis.  
ORGANISM Bacteria: Firmicutes; Bacillus/Clostridium group: Bacillaceae;  
Bacillus.  
1 (bases 1 to 208430)  
AUTHORS Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,  
Azevedo,V., Bertoletti,M.G., Bessieres,P., Bolotin,A., Borchert,S.,  
Boriss,R., Boursier,L., Brans,A., Brum,M., Brignelli,S.C.,  
Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,  
Carter,N.M., Choi,S.K., Codani,J.J., Conerton,I.F., Cummings,N.J.,  
Daniel,R.A., Denicot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,  
Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,  
Fougeron,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,  
Galleron,N., Gilm,S.Y., Glaser,P., Goiteau,A., Goldgilly,E.J.,  
Grandi,G., Guisepi,G., Guy,B.J., Haga,K., Halech,J., Harwood,C.R.,  
Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,  
Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,  
Klaer-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,  
Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,  
Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,  
Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,  
Mizuno,M., Mostl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,  
Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,  
Portetle,D., Portwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,  
Purnelle,B., Rapoport,G., Ray,M., Reynolds,S., Rieger,M.,  
Rivolta,C., Roche,E., Roche,B., Rose,M., Sadale,Y., Sato,T.,  
Scanlan,E., Schleich,S., Schroeder,R., Scoffone,F., Sekiguchi,J.,  
Sekowska,A., Seror,S.J., Serron,P., Shin,B.S., Soldo,B.,  
Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,  
Takeuchi,M., Tamakoshi,A., Tanaka,T., Tepstra,P., Tognoni,K.,  
Tosato,V., Uchiyama,S., Vandendool,M., Vanlier,F., Vassartoli,A.,  
Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weltzenegger,T.,  
Winters,P., Wiputt,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,  
Yoshida,K., Yoshikawa,H.F., Zmistein,E., Yoshikawa,H. and  
Danchin,A.  
TITLE The complete genome sequence of the gram-positive bacterium  
Bacillus subtilis  
JOURNAL Nature 390 (6657), 249-256 (1997)  
MEDLINE 98044033  
REFERENCE 2 (bases 1 to 208430)  
AUTHORS Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.  
TITLE Direct Submission  
JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,

Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724  
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,  
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45  
68 89 48

FEATURES  
source Location/Qualifiers  
1. 208430

organism="Bacillus subtilis"  
strain="168"  
/db\_xref="taxon:1423"  
complement(5..1234)  
/gene="yfnC"  
complement(5..1234)  
/gene="yfnC"  
/function="unknown"  
/note="similar to fosmidmycin resistance protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12561.1"  
/db\_xref="PID:e1182723"  
/db\_xref="PID:g2633056"  
/db\_xref="GI:2633056"  
/db\_xref="SPTREMBL:O06481"

translation="MAIAPLKREKTVQKPGTVVPIILIIIGICHLNDSIAQAVIPAM  
PLERSMSLTFTOLGIAPTLNMSVQPVGYWTDKRPYPALPVGLTASMLGILG  
LAAPSFTIICCFEIGSALFPEGSRYVYAGTRRGIAOSIYOGNSGQAMA  
PIITALLIPLPGFQNAVFTLVAAALAFMTIAKRYASRLSLQKSKQKKTENT  
ATKSVSALIIIFLIFARSWYTSAINFYFVMDYHVSIIQAOSTIEFLLFGA  
IGTFGLGGLADRGKRFVILGSLCSAPLAIPLPAGPVLAVGLALGLVMSFSV  
TVVYAOELVPGKIGTMSGLTGLAFGMGALGAVLALIDAAGLPTMTAIAFLVLG  
ILAFLLPSDOKREWS"  
complement(1345..2052)  
/gene="yfnB"  
complement(1345..2052)  
/gene="yfnB"  
/function="unknown"  
/note="similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12562.1"  
/db\_xref="PID:e1182723"  
/db\_xref="PID:g2633057"  
/db\_xref="GI:2633057"  
/db\_xref="SPTREMBL:O06480"

translation="MKRYRTLLFDVDDPTLIDROAAALATRLIFEDONTPLINDKAO  
KTIHQGLRAPEEKMTREDEVNTRFSALLKEGYEADGALLEKTRRFLKEGQLI  
DGAFLISMLQOQFDLYVTNGVSHQYKRLDSGLFPFKDIFVSEDTGFQPKKEY  
FNHYERLIPQFSAEHTLIIGDSLADIKGQLAGDTCMNDPMKNVPEIIPYEIR  
KLEELYHILIENTVSC"  
complement(2110..3529)  
/gene="yfnA"  
complement(2110..2133)  
/gene="yfnA"  
complement(2144..3529)  
/gene="yfnA"  
/function="unknown"  
/note="similar to metabolite transporter"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12563.1"  
/db\_xref="PID:e1182724"  
/db\_xref="PID:g2633058"  
/db\_xref="GI:2633058"  
/db\_xref="SPTREMBL:O06479"

translation="MSLFRKKPLLETLISQSKSLARTLSAFDILLIGICVIGTGI  
FVITGTAATGAPALIIISFIAGLACALAFCAVAFSSSIPISGSVSYSTVLTGEL  
LAFILGMDLMEYVIALSAVATGSSYROSILAGFNHLIPALITGAPSGMAGAVNLP  
AAVILLITATVSGVKESTRENNVIVLMTAIIILITIVIGYVYKPDNMSPEMGM  
KCVIISATVFRATIGFDVANSASEVKNPQNMFGIISALAVTVLXIANSVLVTC  
NMYPKLVNGDPSFALKFVGQDAVAGIISVAILGITVLMALILYADRLFAFASRD  
GLPGLFAKHSFSTFPFNTLTIGVAGINGFNLGITLALVNMGGLAATVISA  
VIVLRKKHEIKASFSPVPVPIISAGICIMFWISLPGVTWLSFVIMAVGLIYVE  
LYSRKHSLLNK"  
3779..5236

CDS

/gene="yfmR"  
3779..5236  
/gene="yfmR"  
/function="unknown"  
/note="similar to benzaldehyde dehydrogenase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12564.1"  
/db\_xref="PID:e1182725"  
/db\_xref="PID:g2633059"  
/db\_xref="GI:2633059"  
/db\_xref="SPTREMBL:O06478"  
translation="MFOYEELNKOFGIGKMOGSSPNVLEKNPNYTOKTTTPKATA  
DVDBAYRAAALAKKMDAVNPFKRTITLKAATVITIEBERAIIYLIEELGTRIKA  
AFELGLVKNIIKKAATPEIRMEGKILPSTIDGKENRLYRPAVGAVSPENPEFLS  
MKSVAIPALGAGNVILKPHETPIGCGFLIKAFIENAGIPAGLLNVVTVDAIEIGDSF  
VEHPVPRISFTSGTIVKGSYIGOLAMKFKPLLELGSNSAFIYLEDADIEAVAAV  
FSRTHOGQICMSANRVLVHSSITDKLELYQAVESLKVGDPMDDPTIIGPLINSRO  
TDGLMKTVEQALIEGAVPVKLGSGNGTVEPTTIKDVKPEMSIAKEELFGVVSFMR  
DSEDAVDIANETPFGLSGAVHTSLENGVAPAKRIETGMHVDTTINDPENNAFSG  
EKQSGLRNGNSLEFFTLLKWSIVQHKRSFPY"  
5250..6136  
/gene="yfmS"  
5250..6110  
/gene="yfmS"  
/function="unknown"  
/note="similar to methyl-accepting chemotaxis protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12565.1"  
/db\_xref="PID:e1182726"  
/db\_xref="PID:g2633060"  
/db\_xref="GI:2633060"  
/db\_xref="SPTREMBL:O06477"

gene

CDS

terminator

gene

translation="MELTINTEKETADILDAFIKVAPYLSLVODDITIGIYDEKEL  
VNIPAKTFLNWKAGDPLQEGDITDAIRSQKTSWYKELFGFPLIARAPIPHDEN  
GRVIGVGLGTSLSESSKLDHVASLVSQTAASIDISESINGFSTOAGISSOA  
KVSPEAGEINADISVTYVGTSDOSNLGIMAIAPARAGSGSPSVADRIKLTATH  
SKENYQIDQITTKIHSLKGLSEESIESINHTGQAAVQGISATMOEISGSAOHLA  
KMAEKALEEE"  
6113..6136  
/gene="yfmS"  
6245..8159  
/gene="yfmR"  
6245..8134  
/gene="yfmR"

CDS

terminator  
gene  
CDS

/gene="yfmR"  
8138..8159  
/gene="yfmR"  
8257..8740  
/gene="yfmQ"  
8257..8703  
/gene="yfmQ"  
/note="similar to ABC transporter (ATP-binding protein)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12566.1"  
/db\_xref="PID:e1182727"  
/db\_xref="PID:g2633061"  
/db\_xref="GI:2633061"  
/db\_xref="SPTREMBL:O06476"  
translation="MSILKARNLYKTYGDKTLFDHISPHIENRIGLIGNGTKST  
LRKVIAGLESIEGEITKSGSVQVEFLHQDELPAGQVLEHITSGBSAVKTLREYE  
KALYELGDPENEQRKHLAAQAKMANNMNDANTLAKTVLSKGVNDVVRKPVNELS  
GOKKRVAIARNLIOPADLILIDEPNTLIDNETIEWEGEYSQVPGAVMLTHTRYFL  
NRVTNRIYELERGLSYTYKGYEVFLERARERQAQEKTRRONLLRELAMLRGA  
KARSTKOKARIDRYETLKEORTGPOSSGSDPAISHRGKGVIRENNVIMYDGRMIV  
DRFMDVIRPGERIGTIGNGIGKTTLLNALAGRTPPGQDTTIOQYVIRIGYVTDHSE  
MNGELKVIDYIKETAREVKTADGDMITNEQLEFLFPFSKQOTYIRLSGGERRLY  
LLQVLMQEPNVLFDPEPTNDITLSTLEYEIQDFPVVITVSHDRFTLDRGVDRLL  
VFEBSNGVSRQGSYDMESSKAKKAPKRAAEKAEAPKRRKRLSYKDDLEMD  
GIEKIQLEKHHQLEADILAAAGSDFGKIQELMAEQAKTAELEAAMDRTETSLMI  
EELS"  
8138..8159  
/gene="yfmR"  
8257..8740  
/gene="yfmQ"  
8257..8703  
/gene="yfmQ"

/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12567.1"

Query Match 7.8%; Score 17; DB 1; Length 208430;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 TGTCTCTCAGCGCTCCA 68  
|||||  
DB 99303 TGTCTCTCAGCGCTCCA 99287

Search completed: August 6, 1999, 11:38:14  
Job time: 6380 sec

**This Page Blank (uspto)**



Mon Aug 9 13:26:25 1999

us-09-049-696-9.rst

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:04 ; Search time 3075.15 Seconds  
(without alignments)  
139.835 Million cell updates/sec

Title: US-09-049-696-9

Perfect score: 218  
Sequence: 1 GACACATGGCCCTCATGATG.....TCTCTGGATCCCATGAGAC 218

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	156.2	71.7	455	39	AA871197	AA871197 vq32c11.r
2	72.8	33.4	435	33	AA443258	AA443258 aal4d09.r
3	71.8	32.9	430	47	AI493356	AI493356 tg70f11.x
4	53	24.3	468	37	AA726662	AA726662 vu93f05.r
5	47.2	21.7	373	33	AA443218	AA443218 aal4d10.s
6	35.2	16.1	207	41	AF071903	AF071903 AF071903
7	32.8	15.0	362	23	R93192	R93192 yf69g03.s1
8	31.4	14.4	290	20	D22278	D22278 RICC10671A
9	31.4	14.4	327	21	DA7539	DA7539 RICC13108A
10	31.4	14.4	311	31	AA319761	AA319761 EST22025
11	31.2	14.3	389	34	AA457944	AA457944 v174e08.r
12	31.2	14.3	408	47	AI481293	AI481293 v174e09.x
13	31.2	14.3	448	48	AI552284	AI552284 v174e09.y
14	31.2	14.2	501	46	AI439809	AI439809 t156c10.x
15	30.8	14.1	537	43	AI162364	AI162364 A016P43U
16	30.8	14.1	584	43	AI193563	AI193563 g270f08.x
17	30.2	13.9	339	28	C19434	C19434 C19434 Rice
18	30.2	13.9	360	44	AI308681	AI308681 SMOVAFAP
19	30.2	13.9	542	49	AU056068	AU056068 AU056068
20	30.2	13.9	537	49	AU056289	AU056289 AU056289
21	30	13.8	431	28	AA079729	AA079729 zm20e10.r
22	30	13.8	737	39	AA882208	AA882208 vx42b04.r
23	29.6	13.6	505	35	AA584202	AA584202 nol1c05.s
24	29.6	13.6	543	37	AA695481	AA695481 GM02891.5
25	29.6	13.6	400	37	AA698140	AA698140 HL03837.5
26	29.6	13.6	592	38	AA802041	AA802041 GM03766.5
27	29.4	13.5	415	24	H67378	H67378 yu52d06.s1
28	29.4	13.5	372	34	AA506896	AA506896 nh63c04.s
29	29.4	13.5	517	42	AI144770	AI144770 UI-R-BTO
30	29.4	13.5	426	43	AI176877	AI176877 EST220482
31	29.4	13.5	540	46	AI411070	AI411070 EST239364
32	29.4	13.5	600	46	AI411077	AI411077 EST239371
33	29.4	13.5	580	49	AI648253	AI648253 UK37c08.x
34	29.2	13.4	379	43	AI164105	AI164105 A054P23U
35	29.2	13.4	223	43	AI225939	AI225939 UI23f02.y
36	29.2	13.4	572	44	AI272555	AI272555 UK11a02.y
37	29.2	13.4	427	44	AI272555	AI272555 UK11a02.y
38	29	13.3	535	34	AA499559	AA499559 v194g03.r
39	28.8	13.2	400	21	R18556	R18556 yf69g03.s1
40	28.8	13.2	414	30	AA224336	AA224336 z114e02.s
41	28.6	13.1	319	23	H43327	H43327 YP09f04.s1
42	28.6	13.1	420	24	N29367	N29367 yw97b12.s1
43	28.4	13.0	294	22	H11879	H11879 ym15g06.r1
44	28.4	13.0	442	30	AA212884	AA212884 mw78e11.r
45	28.4	13.0	551	31	AA275020	AA275020 vb05d02.r

# ALIGNMENTS

RESULT 1  
AA871197 455 bp mRNA  
LOCUS vq32c11.r1 Barsted bowel MRLB9 Mus musculus cDNA clone  
DEFINITION IMAGE:1095956 5' similar to SW:ECIC\_BOVIN P54281 EPIHELIAL  
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.  
ACCESSION AA871197  
NID g2966642



RESULT 3  
 AI493356/c  
 LOCUS  
 DEFINITION  
 AI493356 430 bp mRNA EST 30-MAR-1999  
 t970f11.x1 Soares.NhMPU.S1 Homo sapiens CDNA clone IMAGE:2114157  
 3' similar to SW:BCIC\_BOVIN P54281 EPITHELIAL CHLORIDE CHANNEL  
 PROTEIN ; mRNA sequence.  
 ACCESSION  
 AI493356  
 NID  
 VERSION  
 AI493356.1 GI:4394359  
 KEYWORDS  
 EST.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 430)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL  
 Unpublished (1997)  
 COMMENT  
 On May 18, 1998 this sequence version replaced gi:3137059.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Insert Length: 396 Std Error: 0.00  
 Seq primer: -40UP from G4bco  
 High quality sequence stop: 367.  
 Location/Qualifiers  
 1..430  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_1lb="IMAGE:2114157"  
 /clone\_1lb="Soares.NhMPU.S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and  
 pregnant uterus"  
 /lab\_host="DH10B"  
 /note="Organ: mixed (see below); Vector: pT713D-Pac  
 (Pharmacia) with a modified polylinker; Site.1: Not I;  
 Site.2: Eco RI. Equal amounts of plasmid DNA from three  
 normalized libraries (melanocyte 2NBM, pregnant uterus  
 NDHPV, and fetal heart NDH19) were mixed, and ss circles  
 were made in vitro. Following BAP purification, this DNA  
 was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from pools of  
 5,000 clones made from the same 3 libraries. The pools  
 consisted of T.M.A.G.E. clones 260232-265223,  
 340488-345479, and 484488-489479."

BASE COUNT 136 a 84 c 80 g 130 t  
 ORIGIN

Query Match 32.9%; Score 71.8; DB 47; Length 430;  
 Best Local Similarity 63.7%; Pred. No. 6.6e-14;  
 Matches 109; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 5 AATGGCCCTATTGATGTTTGGGCCCTTTCATCAGAAATGAGCTGCTCTCAGCGC 64  
 DB 430 AATAGCATGATTTATGCTTTTCAGTAAGATTTCTTGAAGTGAAGCATTTTCAGCAA 371  
 QY 65 TCCATCAGCTTGAGAGTAAGGATTAACCTCCAGAACACCAAGTGAATGAGCACA 124  
 DB 370 CATATTCAGCTTGAAGTACAGGTAAGTAATGTCAAAACCTCATTGTAAGTAAGCACA 311  
 QY 125 GTATATGTCGACACGACCGTGGGAAAGACACTTTGTTTATCACTG 175  
 DB 310 GTGACTGTGATATACCTGTGGCAACGACACTATGTTTCTAGTACGTGG 260

RESULT 4  
 AA726662

LOCUS  
 DEFINITION  
 AA726662 468 bp mRNA EST 02-JAN-1998  
 v093f05.r1 Stratagene mouse skin (#937313) Mus musculus CDNA clone  
 IMAGE:1209729 5' similar to SW:BCIC\_BOVIN P54281 EPITHELIAL  
 CHLORIDE CHANNEL PROTEIN ; mRNA sequence.  
 ACCESSION  
 AA726662  
 NID  
 VERSION  
 AA726662.1 GI:2744369  
 KEYWORDS  
 EST.  
 SOURCE  
 house mouse.  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1 (bases 1 to 468)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Stepien,M., Tan,F., Underwood,R., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE  
 The WashU-HMT Mouse EST Project  
 JOURNAL  
 Unpublished (1996)  
 COMMENT  
 On May 5, 1995 this sequence version replaced gi:798073.  
 Contact: Marra M/Mouse EST Project  
 WashU-HMT Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: -28ml3 rev1 ET from Amersham  
 MGI:646073  
 High quality sequence stop: 443.  
 Location/Qualifiers  
 1..468  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone\_1lb="IMAGE:1209729"  
 /clone\_1lb="Stratagene mouse skin (#937313)"  
 /sex="females"  
 /tissue\_type="whole skin"  
 /dev\_stage="11 weeks old"  
 /note="Organ: skin; Vector: pBluescript SK-; Site.1:  
 EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. Whole skin from 11 week old C57BL/6 female mice.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'  
 adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor  
 sequence: 5' CTCGACTTTTCTTTTCTTTT 3'"

BASE COUNT 121 a 107 c 120 g 120 t  
 ORIGIN

Query Match 24.3%; Score 53; DB 37; Length 468;  
 Best Local Similarity 59.9%; Pred. No. 1.3e-07;  
 Matches 106; Conservative 0; Mismatches 70; Indels 1; Gaps 1;

QY 5 AATGGCCCTATTGATGTTTGGGCCCTTTCATCAGAAATGAGCTGCTCTCAGCGC 64  
 DB 223 AATGGAATGACTGAAGCTTTCGATCTCTTGGAAACAGACATTTTCAGCAA 282  
 QY 65 TCCATCAGCTTGAGAGTAAGGATTAACCTCCAGAACACCAAGTGAATGAGCACA 123  
 DB 283 AGCTTACAGTTGAGAGCGGTGCGAATCTGTGCAACCCAGACCAAGTGGTGTATAC 342  
 QY 124 AGTATGTCGACACGACCGTGGGAAAGACACTTTGTTTATCACTGCAAC 180  
 DB 343 TATGACTGTGATAGCGCGGTGGCAATGACACACTTTTCTAGTACGTGGCAGAC 399

RESULT 5  
 AA726662

AA4443218/C  
LOCUS  
DEFINITION  
ACCESSION  
MID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AA4443218 373 bp mRNA EST 03-JUN-1997  
aa14d10.s1 Soares.NHMP.LSI Homo sapiens cDNA clone IMAGE:813235 3  
similar to TR-G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL  
; mRNA sequence.  
AA4443218  
g2155893  
AA4443218.1 GI:2155893  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 373)  
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gaisel,G., Jost,S.,  
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,  
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B.,  
White,Y., Wyllie,T., Waterston,R. and Wilson,R.  
WashU-Merck BST Project 1997  
unpublished (1997)

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LIDL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information..  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -41m13 fwd. ET from Amersham  
 High quality sequence stop: 359.  
 Location/Qualifiers

**FEATURES**  
**SOURCE**

```

/organism="Homo sapiens"
/db_xref="GDB:604145"
/db_xref="taxon:9606"
/map="188H12; 14q24.3"
/clone="IMAGE:813235"
/clone_1b="Soares_NhMMPy_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHw, pregnant uterus
NbHwU, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

```

	Query Match	21.7%	Score 47.2:	DB 33:	Length 373:
	Best Local Similarity	64.8%:	Pred. No. 1e-05:	Mismatches 38:	Indels 0; Gaps 0
	Matches	70:	Conservative	0:	
Qy	68 ATCCAGCTTAGAAGTAAAGGATTAAACCCCTCCAGAACGCCAGTGATGATGCACAGTG	127			
Db	367 ATTCAAGCTTGAAGAATNACAGGTGAATAATGTCNAACCTCACACATCAATTGAAAAACAAGTG	308			
Qy	128 ATCTGGACAGCACCCTGGGAAAGACACTTTGTTCTTATACCTGG	175			
Db	307 ACTGTGATTAATACTGTGTGGGCAACGCACTAATGTTTCTAGTTACCTGG	260			
RESULT	6				
AF071903/c	*				

LOCUS AF071903 207 bp mRNA EST 30-JUN-1998  
DEFINITION AF071903 Fetal aorta endothelial cell line CFA47 Bos taurus CDNA  
ACCESSION AF071903  
NID 93265169  
VERSION AF071903.1 GI:3265169  
KEYWORDS EST.  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (Bases 1 to 207)  
AUTHORS Oberts,T.B. and Alliegro,M.C.  
TITLE P1gpen-binding mRNA  
JOURNAL Unpublished (1998)  
COMMENT On May 18, 1995 this sequence version replaced gi:811149.

Contact: Mark C. Allegro  
Anatomy  
Louisiana State University Medical Center  
1901 Perdido Street, New Orleans, LA 70112, USA  
Email: mallee@lsu.mc.edu.

FEATURES	Location/Qualifiers
source	1. .207

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/map="1"
/clone="38"
/clone_lib="Fetal aorta endothelial cell line CPA47"
/issue_type="aorta"
/cell_type="endothelium"
/cell_line="CPA47"
/dev_stage="fetus"
/note="pigpen (GenBank Accession Number U26024) -binding
mRNA"

```

Query Match	16.1%;	Score 35.2;	DB 41;	Length 207;
Best Local Similarity	62.5%;	Pred. No. 0.082;		
Matches 55;	Conservative 0;	Mismatches 33;	Indels 0;	Gaps 0;

[illegible]

RESULT	7
LOCUS	R93192
DEFINITION	362 bp mRNA EST 22-JAN-1999
ACCESSION	U66960.3 s1 Soares fetal liver spleen INTLS Homo sapiens cDNA clone IMAGE:275884 3', mRNA sequence.
VERSION	R93192.1
KEYWORDS	GI:967358
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 362) Hillier, L., Clark, N., Dubucq, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennson, G., Marra, M., Parsons, J., Rifkin, D., Rohlfing, T., Soares, M., Tan, E., Tsvetaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE	The Washo-Merck EST Project



/db\_xref="taxon:4530"  
/clone\_lib="Rice green shoot"  
/note="Green shoot (8 days old)"  
BASE COUNT 64 a 102 c 101 g 59 t 1 others  
ORIGIN

Query Match 14.4% Score 31.4; DB 21; Length 327;  
Best Local Similarity 56.2%; Pred. No. 1.8;  
Matches 59; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 39 CAGGAATGAGACTCTCTCTCAGCCCTCATCCACTTGAGAGTAAAGGATTAACCTTCC 98  
Db 218 CGGAACAGCGCGTGTCTTGACGAGCTCCACACTCCGGCAGAGAATCTTCACCTTCC 159  
OY 99 AGAACAGCCAGTGAATGATGACAGTGAATGATGACAGCAGCAGC 143  
Db 158 TGAACGCCGCCGGGATGAACGCCGCCGCTTCGGGTTCAAGCAGC 114

RESULT 10  
AA319761/c  
LOCUS AA319761 311 bp mRNA EST 19-APR-1997  
DEFINITION EST22025 Adrenal gland tumor Homo sapiens CDNA 5' end. mRNA  
SEQUENCE  
AA319761  
NID g1972087  
VERSION AA319761.1 GI:1972087  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
AUTHORS  
1 (bases 1 to 311)  
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C.,  
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fline,L.D.,  
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Georgagen,N.S.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,  
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,  
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pellegriano,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shiley,R.,  
Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,  
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dumke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
He,M.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M. and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
66026280  
On Sep 12, 1996 this sequence version replaced gi:1394061.

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423

FEATURES  
SOURCE  
1. .311  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):120318"  
/db\_xref="taxon:9606"  
Email: atkerlavet@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tadb/ngi/ngi.html>)  
Seq primer: M13 Reverse.

/map="8"  
/clone\_lib="Adrenal gland tumor"  
/dev\_stage="adult"  
/note="Organ: adrenal gland; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 65 a 87 c 83 g 74 t 2 others  
ORIGIN

Query Match 14.4% Score 31.4; DB 31; Length 311;  
Best Local Similarity 61.7%; Pred. No. 1.8;  
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY 73 GCTTGAAGTAAAGGATTAACCTTCAGAACAGCCAGTGAATGACAGTATGT 132  
Db 310 GTTGGGGCTTAGGGCTTAGTCATCAGAGACAGTGAAGGCGGATGCTGCTT 251  
OY 133 GGACAGCACCGCTGGGAAAGCA 153  
Db 250 CGTCAGTCTGCTGTGAAGCA 230

RESULT 11  
AA457944/c  
LOCUS AA457944 389 bp mRNA EST 06-JUN-1997  
DEFINITION V174609.F1 Soares mouse mammary gland NbMNG Mus musculus CDNA clone  
IMAGE:849544 5', mRNA sequence.  
AA457944  
NID g2180664  
VERSION AA457944.1 GI:2180664  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
1 (bases 1 to 389)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Gaisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellhammer,K., Stepec,M., Tan,F., Underwood,K., Moore,B.,  
Raisling,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and  
Westerberg,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Sep 12, 1996 this sequence version replaced gi:1292013.

TITLE  
JOURNAL  
COMMENT  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
MGI:501696  
Seq primer: -28m13 rev2 ET from Amerisham  
High quality sequence stop: 385.  
Location/Qualifiers  
1. .389

FEATURES  
SOURCE  
1. .389  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/map="1, 14; 1, 14"  
/clone\_image="IMAGE:849544"  
/clone\_lib="Soares mouse mammary gland NbMNG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland; Vector: pT73D-Pac  
(Pharacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5'

TTGACCAATCTGAGTGGAGCGCGCGCAATGTTTTTTTTTTTTTTTTTTT  
 T31; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M. Fatima  
 Bonaldo."

BASE COUNT 111 a 81 c 96 g 101 t  
 ORIGIN

Query Match 14.3% Score 31.2; DB 34; Length 389;  
 Best Local Similarity 54.3%; Pred. No. 2.3;  
 Matches 63; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 15 TTGATGCTTTGGGCGCTTCATCAGAAATGAGCTGTCTCAGCCCTCCATCCAGC 74  
 DB 256 TTGATCTCATGAGGAGCATTTCTCACTGAGAGCTCTTCTGTATTAATCTCAGCTGAG 197  
 OY 75 TTGAGAGTAAGGAGATTACCTCCAGACAGCCAGTGTGATGAGTGCACAGTATC 130  
 DB 196 TCAAGTTGACACAAACACAGCCAGTGTGATCTGTGAGCCACATGAGC 141

RESULT 12  
 A1481293 408 bp mRNA EST 09-MAR-1999  
 LOCUS v174e09.x1 Soares mouse mammary gland NDMMG Mus musculus cDNA clone  
 DEFINITION IMAGE:849544 3', mRNA sequence.  
 A1481293  
 ACCESSION G4374519  
 NID A1481293.1 GI:4374519  
 VERSION EST.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 408)  
 AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R. and Wilson, R.  
 TITLE The WashU-NCI Mouse EST Project 1999  
 JOURNAL Unpublished (1999)  
 COMMENT On Jun 22, 1998 this sequence version replaced gi:3247231.

CONTACT: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:501696  
 This clone was previously sequenced on the 5' end only, this new  
 data is from the 3' end  
 High quality sequence stop: 351.  
 Location/Qualifiers  
 1..408  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:849544"  
 /clone\_lib="Soares mouse mammary gland NDMMG"  
 /sex="male"  
 /tissue\_type="mammary gland"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Organ: mammary gland; Vector: pT73D-Pac  
 (Pharmacia) with a modified polylinker; Site 1: Not I;  
 Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5',  
 TGTACCAATCTGAGTGGAGCGCGCGCAATGTTTTTTTTTTTTTTTTTTT  
 T31; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M. Fatima  
 Bonaldo."

BASE COUNT 98 a 98 c 89 g 123 t  
 ORIGIN

Query Match 14.3% Score 31.2; DB 47; Length 408;  
 Best Local Similarity 54.3%; Pred. No. 2.3;  
 Matches 63; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 15 TTGATGCTTTGGGCGCTTCATCAGAAATGAGCTGTCTCAGCCCTCCATCCAGC 74  
 DB 273 TTGATCTCATGAGGAGCATTTCTCACTGAGAGCTCTTCTGTATTAATCTCAGCTGAG 332  
 OY 75 TTGAGAGTAAGGAGATTACCTCCAGACAGCCAGTGTGATGAGTGCACAGTATC 130  
 DB 333 TCAAGTTGACACAAACACAGCCAGTGTGATCTGTGAGCCACATGAGC 388

RESULT 13  
 A1552284 448 bp mRNA EST 23-MAR-1999  
 LOCUS v174e09.y1 Soares mouse mammary gland NDMMG Mus musculus cDNA clone  
 DEFINITION IMAGE:849544 5', mRNA sequence.  
 A1552284  
 ACCESSION G4484647  
 NID A1552284.1 GI:4484647  
 VERSION EST.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 448)  
 AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R. and Wilson, R.  
 TITLE The WashU-NCI Mouse EST Project 1999  
 JOURNAL Unpublished (1999)  
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2948905.

CONTACT: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:501696  
 This read is a RESEQUENCE of a previously sequenced mouse clone  
 This read has been verified (found to hit its original self in the  
 correct orientation)  
 Seq primer: -40RP from G1bco  
 High quality sequence stop: 447.  
 Location/Qualifiers  
 1..448  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /map="12"  
 /clone="IMAGE:849544"  
 /clone\_lib="Soares mouse mammary gland NDMMG"  
 /sex="male"  
 /tissue\_type="mammary gland"  
 /dev\_stage="4 weeks"







**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:03:09 ; Search time 921.96 Seconds  
(without alignments)  
59.159 Million cell updates/sec

Title: US-09-049-696-9

Perfect score: 218  
Sequence: 1 GAACATGGCCTCATTGATG.....TCTCTGGATCCCAATGAC 218

Scoring table: IDENTITY\_MUC

Searched: 311585 seqs, 125096042 residues

Database : N.Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29.8	13.7	5455	1 V33577	Osteotesticular pr
2	29.4	13.5	464	1 Q20988	CDR-grafted, human
3	29.4	13.5	464	1 Q27355	Encodes ASB7 GH-2
4	29.4	13.5	777	1 T42509	Humanised ASB7 Fd
5	28.6	13.1	7812	1 X24753	Human interleukin-
6	28	12.8	359	1 Q46579	Active site polynu
7	27.8	12.8	2724	1 Q12226	N-alpha-acetyltan
8	27.6	12.7	464	1 Q20987	CDR-grafted, human
9	27.6	12.7	464	1 Q27354	Encodes ASB7 GH-1
10	27.4	12.6	150	1 X10506	Human biallelic po
11	27.4	12.6	769	1 T98688	DNA encoding a S.
12	27.4	12.6	769	1 X30779	Streptococcus pneu
13	26.8	12.3	787	1 Q27946	Sequence encoding
14	26.8	12.3	423	1 Q75094	Plasmid pOKSc4 co
15	26.8	12.3	1088	1 V34275	Human secreted pro
16	26.8	12.3	1143	1 V34222	Human secreted pro
17	26.4	12.1	28804	1 T37329	Sphingomonas genus
18	26.4	12.1	28804	1 T92474	Sphingomonas genus
19	26.4	12.1	28804	1 V81474	Chromosomal fragme
20	26.4	12.1	477	1 V68416	CDNA encoding TNF2
21	26.4	12.1	28804	1 V99812	Sphingomonas S88 s
22	26.2	12.0	2277	1 Q89338	Mouse STAT1 CDNA.
23	26.2	12.0	2277	1 T31278	Mouse STAT1 CDNA.
24	26.2	12.0	3496	1 T48237	Neuronal nicotinic
25	26.2	12.0	2082	1 T59527	Alpha4 subunit of
26	26.2	12.0	2082	1 T59528	Alpha4 subunit of
27	26.2	12.0	2363	1 V12196	Human neuronal nic
28	26	11.9	3955	1 V15181	Human seriate 2 en
29	25.6	11.7	3568	1 T28638	RNase L inhibitor
30	25.6	11.7	7684	1 X12953	Enterococcus faeca
31	25.4	11.7	4394	1 Q21604	Alpha galactosidas
32	25.4	11.7	388	1 T35132	Diminished express
33	25.4	11.7	651	1 V01702	Pleurocoid growt
34	25.4	11.7	11832	1 N80863	Nucleotide sequen
35	25.2	11.6	2776	1 N80863	Sequence of CDNA c
36	25.2	11.6	2776	1 Q06050	Sequence encoding
37	25.2	11.6	2776	1 Q22780	Codes for beta-sub
38	25.2	11.6	1565	1 T02334	GABA-A receptor ga
39	25.2	11.6	2310	1 V64854	Human LFA-1 beta c
40	25.2	11.6	32768	1 X13060	Enterococcus faeca
41	25.2	11.5	348	1 X00880	Mouse derived RT3
42	25	11.5	8584	1 T35166	Plasmid pHTL-D2 (M
43	25	11.5	8590	1 T35168	Plasmid pD2pick (MF

## ALIGNMENTS

C	44	25	11.5	8157	1	T35165
	45	25	11.5	110000	1	T58840_2
RESULT 1						
DR	V33577	standard; DNA; 5455 BP.				
AC	V33577;					
DT	29-DEC-1998	(first entry)				
DE	Osteotesticular protein tyrosine phosphatase cDNA.					
KW	Osteotesticular protein tyrosine phosphatase; OST-PTP; bone; testis;					
KW	osteoblast specific form; OST; OST-PTP receptor; osteoporosis;					
OS	Rattus sp.					
FH	Key	Location/Qualifiers				
FT	CDS	205..5340				
FT		/tag= a				
FT		/product= "OST-PTP"				
FT		5449				
PN	polyA_signal	/tag= b				
PD	US5821084-A.					
PE	13-OCT-1998.					
PR	21-NOV-1994; 342930.					
PA	(AMHP ) AMERICAN HOME PROD CORP.					
PA	(UNMI ) UNIV MICHIGAN.					
PI	Davis AR, Dixon JE, Mauro LJ, Olmsted EA;					
DR	WPI: 98-567659/48.					
DR	P-PSDB: W70506.					
PT	DNA encoding osteoblast-testicular protein tyrosine kinase					
PT	polypeptide - useful for producing recombinant polypeptide					
PS	Claim 1: Columns 23-24; 31pp; English.					
CC	The present sequence represents a rat osteotesticular protein tyrosine					
CC	phosphatase (OST-PTP) cDNA which was isolated from the fetal rat					
CC	calvaria and the rat osteosarcoma cell line, UMR106. OST-PTP is a					
CC	transmembrane protein expressed only in bone and testis. The OST-PTP					
CC	CDNA is useful for producing recombinant protein. The invention also					
CC	provides a truncated osteoblast specific form (OST) containing the					
CC	OST-PTP receptor but lacking the catalytic domain. The OST-PTP protein					
CC	is claimed to be useful in assays used for screening abnormal bone					
CC	growth patterns and metabolic bone diseases. The OST-PTP protein is					
CC	also claimed to be useful in the treatment of osteoporosis, osteopetrosis					
CC	and other bone metabolic disorders.					
SO	Sequence 5455 BP; 1169 A; 1607 C; 1525 G; 1154 T;					
Query Match						
Best Local Similarity 13.7%; Score 29.8; DB 1; Length 5455;						
Matches 70; Conservative 0; Mismatches 67; Indels 0; Gaps 0;						
QY	46	TGAGTGTCTCTGACGCTTCATCCAGTGTGAGTAGGATTAACCTCCAGAACAG	105			
DB	5049	TGATGTGTGGCCATGCTCTCCGGGCAAGACCAAGCCAGGACCTGCTCAG	5108			
QY	106	CCAGTGTGATGATGACAGATGATCTGTGACAGACACCGTGGAAAGACACTTGTTCCT	165			
DB	5109	CCATCTCAGCAAAAACAAACACAGCTGGACCTTCTGGCATGACAGAGCTGTATA	5168			
QY	166	TATCACTGTGACACGC	182			
DB	5169	GCAAGCAGGGACAGACG	5185			
RESULT 2						
ID	Q20988	standard; DNA; 464 BP.				
AC	Q20988;					
DT	19-MAY-1992	(first entry)				
DE	CDR-grafted, humanised heavy chain GH2.					
KW	murine monoclonal antibody; Mab; ASB7; humanised antibody; CEA;					

KW Complementarity determining region: ss.  
OS Homo sapiens.  
OS Mus musculus.  
FH Key Location/Qualifiers  
FT signal\_peptide 22..78  
FT /\*tag= a  
FN WO9201059-A.  
PD 23-JUN-1992.  
PF 05-JUL-1991; G01108.  
PR 05-JUL-1990; GB-014932.  
PR 21-DEC-1990; WO-G02017.  
PR 05-JUL-1991; WO-G01108.  
PA (CELL-) CELTECH LTD.  
PI Adair JR, Bodmer MW, Mountain A, Owens RJ;  
PI WPI: 92-056874/07.  
DR P-PSDB: R2373.  
PT New CDR-grafted anti carcinoembryonic antigen antibodies - useful  
PT in therapy and diagnosis of carcinoma  
PS Example 4; Fig 11; 70pp; English.  
CC This heavy chain sequence comprises a human framework (i.e. the LAY  
CC region) which contains murine sequences (from the murine anti-CEA  
CC A5b7 Mab) in the CDRs and at other positions predicted to be  
CC important for antigen-binding and at which human and A5b7 sequences  
CC differ. The GH2 heavy chain encoded by this sequence has murine  
CC CDRs at residues 26-35 (CDR1), 50-65 (CDR2) and 95-102 (CDR3) and  
CC additional murine residues within the framework at residues 1, 48,  
CC 49, 72, 73, 76, 82b, 86 and 93. Numbers apply to the mature variable  
CC region. (See Q020984 for A5b7 heavy chain coding sequence).  
SQ Sequence 464 BP; 122 A; 110 C; 130 G; 102 T;

Query Match 13.5%; Score 29.4; DB 1; Length 464;  
Best Local Similarity 49.7%; Pred. No. 1.3;  
Matches 75; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

OY 36 CATCAGAAATGAGACTGTCTCTCAGCGCTCCATCCAGCTTGAGATGAAGGATTAACC 95  
DB 197 CACCTGGAAAGGACTCGATGCTGGCTTTCGAAATAGCAATGATGATACACAA 256  
OY 96 TCAGAACAGCCAGTGTGATGACAGTGTGACACCGTGGAGAAAGACA 155  
DB 257 CAGAGTACTGTGATCTGTGAAGGAGATTCACAAATTCAGACAAAGACAGTCA 316  
OY 156 CTTGTCTTATCACTCGGACACGACGCC 186  
DB 317 CACTGTACTCGACATGATACACTGCAGGC 347

RESULT 3  
Q27355 ID Q27355 standard; DNM: 464 BP.  
AC Q27355;  
DT 03-FEB-1993 (first entry)  
DE Encodes A5b7 GH-2 antibody grafted heavy chain.  
KW humanised antibody; chimeric; carcino-embryonic antigen; therapy;  
KW diagnosis; carcinomas; CDR; 19G; human; murine; ss.  
OS Chimeric.  
FH Key Location/Qualifiers  
FT cds 22..457  
FT /\*tag= a  
FT /\*tag= b  
FT /\*tag= "murine grafted CDR1"  
FT /\*tag= "murine grafted CDR2"  
FT /\*tag= "murine grafted CDR3"  
FT /\*tag= d  
FT /\*tag= "murine grafted CDR3"  
PN WO9201059-A.  
PD 23-JUN-1992.  
PF 05-JUL-1991; G01108.  
PR 05-JUL-1990; GB-014932.

PR 21-DEC-1990; WO-G02017.  
PR 05-JUL-1991; WO-G01108.  
PA (CELL-) CELTECH LTD.  
PI Adair JR, Bodmer MW, Mountain A, Owens RJ;  
PI WPI: 92-284316/34.  
DR P-PSDB: R26151.  
PT Humanised antibody molecules - comprising murine and human regions,  
PT specific for carcino-embryonic antigen, useful for diagnosis and  
PT therapy.  
PS Example 4; Figure 8; 71pp; English.  
CC This sequence encodes a CDR-grafted A5b7 human antibody heavy chain  
CC having murine CDRs at amino acids 26-35 (CDR1), 50-65 (CDR2) and  
CC 95-102 (CDR3) and additional murine framework residues at 1, 48, 49, 72,  
CC 73, 76, 82b and 86. The LAY framework was chosen when making the  
CC construct as it shows the highest homology to A5b7.  
SQ Sequence 464 BP; 122 A; 110 C; 130 G; 102 T;

Query Match 13.5%; Score 29.4; DB 1; Length 464;  
Best Local Similarity 49.7%; Pred. No. 1.3;  
Matches 75; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

OY 36 CATCAGAAATGAGACTGTCTCTCAGCGCTCCATCCAGCTTGAGATGAAGGATTAACC 95  
DB 197 CACCTGGAAAGGACTCGATGCTGGCTTTCGAAATAGCAATGATGATACACAA 256  
OY 96 TCAGAACAGCCAGTGTGATGACAGTGTGACACCGTGGAGAAAGACA 155  
DB 257 CAGAGTACTGTGATCTGTGAAGGAGATTCACAAATTCAGACAAAGACAGTCA 316  
OY 156 CTTGTCTTATCACTCGGACACGACGCC 186  
DB 317 CACTGTACTCGACATGATACACTGCAGGC 347

RESULT 4  
T42509 ID T42509 standard; DNM: 777 BP.  
AC T42509;  
DT 17-FEB-1997 (first entry)  
DE Humanised A5b7 Fd fragment gene.  
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;  
KW mustard-ribonucleotide; antibody directed enzyme product therapy;  
KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;  
KW reduced immunogenicity; non-selective triggering; primer;  
KW polymerase chain reaction; PCR; HP-RNase; Fd; F(ab')2; ss.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT cds 16..768  
FT /\*tag= a  
FN WO9620011-A1.  
PD 04-JUL-1996.  
PF 21-DEC-1995; G02991.  
PR 23-DEC-1994; GB-026192.  
PR 16-AUG-1995; GB-016810.  
PA (ZENNE) ZENNECA LTD.  
PI Blakey DC, Boyle FT, Davies DH, Eggelte RJ, Heaton DW;  
PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;  
PI Farragana-Fiol A, Taylorson CJ;  
PI WPI: 96-321650/32.  
DR P-PSDB: W06179.  
PT Two component system for anti-tumour therapy - comprising targeting  
PT moiety linked to mutated enzyme which can transform an  
PT anti-neoplastic prodrug  
PS Example 6; Page 121-122; 182pp; English.  
CC A two-component system for anti-tumour therapy comprises a targeting  
CC moiety linked to a mutated enzyme which can transform an anti-neoplastic  
CC prodrug. The system is based on antibody directed enzyme product therapy  
CC (ADEPT) using a non-naturally occurring mutant form of a host enzyme,  
CC (pref. human pancreatic ribonuclease (HP-RNase), (see T42478-83). The  
CC targeting moiety can be an antibody, in partic. murine monoclonal  
CC antibody A5b7 (which binds to human carcinoembryonic antigen) A5b7 is  
CC suitable for targeting colorectal carcinoma. Fragments, esp. F(ab')2,

CC of the antibody can be conjugated to HP-RNase A5B7 Fd and L chain  
CC fragments were isolated by PCR using cDNA isolated from A5B7 hybridoma  
CC cells. The present sequence encodes a humanised A5B7 Fd fragment.  
50 Sequence 777 BP; 193 A; 215 C; 207 G; 162 T;

Query Match	13.5%	Score 29.4	DB 1	Length 777
Best Local Similarity	49.7%	Pred. No. 1.6		
Matches 75	Conservative 0	Mismatches 76	Indels 0	Gaps 0

QY 36 CATAGGAAATGGAGGCTGTCTCTCAGGGCTCATACAGCTTGAAGATTAAGGGATTAACCC 95  
 Db 191 CACCTGGAAAGGAGCTCGAGTGGCTTGGGCTTCATCTGGAATTAAGGCAAAATGACACCA 250  
 QY 96 TCCAGAACAGCCAACTGGATGATGGCAGATGATCTGGACAGCAGCCGTGGGAAAGACA 155  
 Db 251 CAGAGTACTCTGCACTCTGTGTGAAGGAAAGATTACAAATTTCCAGAGACAAAGCAAGTCCA 310  
 QY 156 CTTGTCTTATCACTCGACACGAGCC 186  
 Db 311 CACTGTACTTCAGATGATACACTGCAAGC 341

RESULT 5  
 ID X24753 standard. CDNA: 7812 BP.  
 AC X24753;  
 DT 21-JUN-1999 (first entry)  
 DE Human interleukin-18 binding protein gene.  
 KW Interleukin-18 binding protein; IL-18BP; IL-18BPb; splice variant;  
 KW human; autoimmune disease; inflammation; diabetes; pancreatitis;  
 KW rheumatoid arthritis; graft rejection; sepsis; chronic hepatitis;  
 KW peristalsis; inflammatory bowel disease; multiple sclerosis;  
 KW ischemic heart disease; ischemic brain injury; gene therapy; ss.  
 OS Homo sapiens.  
 PN MO9909063-A1.  
 PD 25-FEB-1999.  
 PF 13-AUG-1998: IL0379.  
 PR 22-JUL-1998: IL-125463.  
 PR 14-AUG-1997: IL-121554.  
 PR 27-AUG-1997: IL-121639.  
 PR 29-SEP-1997: IL-121860.  
 PR 06-NOV-1997: IL-122134.  
 PA (YEDA ) YEDA RES & DEV CO LTD.  
 PI Dinarello C, Kim SH, Novack D, Rubinstein M;  
 WP: 99-180975/15.  
 PT New interleukin-18 binding protein - useful for treating human  
 diseases, including autoimmune disease and inflammation  
 PS Disclosure: Page 60-63; 100pp: English.  
 CC The present sequence comprises human interleukin-18 binding protein  
 CC (IL-18BP) genomic DNA. It was produced from 5 genomic clones  
 CC isolated from a human genomic library by screening with full-length  
 CC human IL-18BP cDNA. No exon coding for a transmembrane receptor  
 CC was identified within the 7.8 kb sequence. 4 Splice variant clones  
 CC (see X24749-52) were identified that code for putative soluble  
 CC secreted proteins, designated IL-18BPa-d (see W98004-007). The  
 CC IL-18BP locus contains an additional gene, coding for the nuclear  
 CC mitotic apparatus protein 1 (NUMA1), positioned on the minus strand  
 CC than localised the IL-18BP gene to human chromosome 11q13(36).  
 CC IL-18BP polypeptides capable of binding IL-18 and/or modulating  
 CC and/or blocking IL-18 activity are provided by the invention.  
 CC Methods for their isolation and recombinant production, DNA vectors  
 CC expressing them, vectors useful for their expression in humans and  
 CC other mammals, and antibodies against them are also provided.  
 CC IL-18BP polypeptides, and DNA encoding them, can be used to treat  
 CC conditions requiring the protein (claimed). Such conditions  
 CC include autoimmune diseases, type I diabetes, rheumatoid arthritis,  
 CC graft rejections, inflammatory bowel disease, sepsis, multiple  
 CC sclerosis, ischemic heart diseases, ischemic brain injury, chronic  
 CC hepatitis, peristalsis, and chronic/acute pancreatitis. IL-18BP is  
 CC also useful for purifying IL-18 (claimed).  
 SQ Sequence 7812 BP; 1689 A; 2278 C; 2167 G; 1678 T;

Query Match	13.1%;	Score 28.6;	DB 1;	Length 7812;
Best Local Similarity	57.1%;	Pred. No. 6.7;		
Matches 52;	Conservative	0;	Mismatches 39;	Indels 0;
				Gaps 0;

Oy	16	TGAGCTTTTGGGGCCCTTTTCATCAGAAATGGAGCTGTCTCATACGCCTCAATCAACT	75
Db	1946	TGCTGAAGTTGGGGAAAGCGGCTGCAGAGCCACACAGGATAAAGTCAAGCGTTCATTAGG	1887
Oy	76	TGAGACTTAAGGAGTAAACCTCCAGAACC	106
Db	1886	ACACAGCACAGAGTTAGCTTTACAGAGAAC	1856

RESULT 6  
ID 046579/c  
AC 046579 standard; DNA; 359 BP.  
DT 22-DEC-1993 (first entry)  
DE Active site polynucleotide fragment.  
KW Active site; active region; prediction; secondary structure;  
KW energy value; engineering; ss.  
OS Synthetic.  
PN J05155899-A.  
PD 22-JUN-1993  
PF 23-AUG-1991  
PF 23-AUG-1991; JP-212284.  
PA (SAGA ) SAGAMI CHEM RES CENTRE.  
DR WPI; 93-232353/29.  
PT Prediction of the active site in physiologically active  
PT polypeptide - by determ. of sec. structure energy values of  
PT partial regions along prim. sequence of physiologically active  
PT polynucleotide, etc.  
PS Disclosure: Page 23-24: 43pp; Japanese.  
CC This sequence represents a polynucleotide which was used to  
CC demonstrate the method of the invention. This oligomer represents  
CC an active site/region and the method of the invention may be used to  
CC predict the active site of a polynucleotide by determining the energy  
CC values of the secondary structure of various partial regions of the  
CC polynucleotide to be tested and the primary sequence as well. At  
CC least one partial region should show a local peak energy value  
CC significantly higher or lower than the average standard energy value.  
CC This method may be used in the elucidation of the mechanism of  
CC action of polynucleotides or polypeptides, and in the engineered  
CC improvement of such sequences.  
QZ Sequence 359 BP; 69 A; 87 C; 110 G; 93 T;

Query Match	12.8%;	Score 28;	DB 1;	Length 359;
Best Local Similarity	51.8%;	Pred. No. 3.5;		
Matches	87;	Mismatches	80;	Indels 1;
				Gaps 1

QY	29	GCCCTTCATCAGGAAATGAGAGCTGTCTCAGCGCTCCATCCAGCTTGAGAGTAAAGGA	88
Db	322	GGCCGCGATAGTGTAGAGGGGGGGCTCTCAAACTTCATGTGATGAAATACCCG	263
QY	89	TTAACCTCCGAAACACCCAGTATGTAATATGCGACATGATCTGGGACACACCGTGGGA	148
Db	262	TCCGCGCTAATGCGAGATCCGGGAACCATGACCAACCAACAAAGAGATCAACCCAGAGA	203
QY	149	AAGGACACTTGTCTTCTTATCACTGTGACAAACGACGCTCCCAATC	196
Db	202	AA-CACACGTTGTGTAATTATTAACCTGGTACGCTTGATAGCTACAC	156

RESULT 7  
Q12226/c  
ID Q12226 standard; cDNA, 2724 BP.  
AC Q12226;  
DT 02-AUG-1991 (first entry)  
DE N-alpha-acetyltransferase AA1 gene.  
KW N-alpha-acetyltransferase; amino acid sequencing; AA1 gene; ss  
OS Saccharomyces cerevisiae.

```

FH Key Location/Qualifiers
FT cds 22.2583
FT /*tag= a
FT /*product= N-alpha-acetyltransferase
PN W09106673-A.
PD 16-MAY-1991.
PF 15-OCT-1990; U056883.
PR 25-OCT-1989; US-426381.
PA (GCHO-) GEN HOSPITAL CORP.
PI Smith JA, Lee FJS.
DR WPI: 91-164219/22.
DR P-PSDB: R12042.
FT Mutant N-alpha-acetyl-transferase - produced from Saccharomyces
FT cerevisiae for use in amino acid sequence determ.
PS Disclosure: Fig 1; 77pp; English.
CC The AAI gene is located on chromosome IV and is positioned
CC adjacent to the 5' flanking sequence of the SIR2 gene.
CC Cells contg. a mutated AAI gene lack N-alpha-acetyltransferase
CC activity and are used to express, in vitro a recombinant protein or
CC peptide lacking an acetyl gp. at the alpha-amino gp. or to produce
CC heterologous proteins. The proteins produced have altered
CC N-alpha-acetylation characteristics, e.g. increased or decreased
CC substrate specificity and thermal stability. The amino acid
CC sequence of such proteins and peptides can be sequenced.
SQ Sequence 2724 BP; 953 A; 491 C; 533 G; 747 T;

Query Match 12.8%; Score 27.8; DB 1; Length 2724;
Best Local Similarity 49.0%; Pred. No. 8.5;
Matches 74; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

OY 19 TCGTTTGGGCGCTTATCATCGAATGAGACTGCTCTACGCGCTCATCAGCTTGA 78
DB 832 TTTCTTTGATGACAGTTTATTAACATGAGCGCGCTTCAATGACCAATTCATGT 773
OY 79 GAGTAGGAGATTACCCCTCCAGACAGCCAGTGAATGATGACAGATGATGAGCAG 138
DB 772 AATATGTTGCTTTCTCTCTATTAACCAATTATCAAGACGATGCTGATATCAT 713
OY 139 CACCGTGGGAAGACACTTGTCTTATC 169
DB 712 TCAATGTTCAATACATTTGTACTTGTCT 682

RESULT 8
OY 8
ID Q20987 standard; DNA; 464 BP.
AC Q20987;
DE 19-MAY-1992 (first entry)
DE CDR-grafted, humanised heavy chain gH1.
KM murine monoclonal antibody; Mab; A5B7; humanised antibody; CEA;
KM complementarity determining region; ss.
OS Homo sapiens.
OS Mus musculus.
FH Key Location/Qualifiers
FT signal_peptide 22.78
FT /*tag= a
FT W09201059-A.
PN 23-JAN-1992.
PF 05-JUL-1991; G01108.
PR 05-JUL-1990; GB-014932.
PR 21-DEC-1990; WO-602017.
PR 05-JUL-1991; WO-G01108.
PA (CELL-) CELLTRECH LTD.
PI Adair JR, Bodmer MW, Mountain A, Owens RJ;
DR WPI: 92-056874/07.
DR P-PSDB: R20793.
FT New CDR-grafted anti carcinoembryonic antigen antibodies - useful
FT in therapy and diagnosis of carcinoma
PS Example 4; Fig 10; 70pp; English.
CC This heavy chain sequence comprises a human framework (i.e. the LAY
CC region) which contains murine sequences (from the murine anti-CEA
CC A5B7 Mab) in the CDRs and at other positions predicted to be
```

```

CC important for antigen-binding and at which human and A5B7 sequences
CC differ. The gH1 heavy chain encoded by this sequence has murine
CC CDRs at residues 26-35 (CDR1), 50-65 (CDR2) and 95-102 (CDR3) and
CC additional murine residues within the framework at residues 1, 48,
CC 49, 72, 73, 76 and 93. Numbers apply to the mature variable region.
CC (See Q20984 for A5B7 heavy chain coding sequence).
SQ Sequence 464 BP; 120 A; 108 C; 133 G; 103 T;

Query Match 12.7%; Score 27.6; DB 1; Length 464;
Best Local Similarity 50.8%; Pred. No. 5.2;
Matches 66; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

OY 36 CATCAGGAATGAGACTTCTCTACGCGCTCATCAGCTTGAAGGATTAAGCC 95
DB 197 CACCTGGAAAGGACTCAGAGGCTGGGCTTATCGAATAGCAATGATACACAA 256
OY 96 TCCAGAACAGCAGATGATGATGACAGTATGTCGACAGCAGCGTGGAAAGACA 155
DB 257 CAGAGTACTCTGCATCTGTGAAGGAGATTCACATTTCCAGAGACAGCAAGTCCA 316
OY 156 CTGTGTTTCT 165
DB 317 CACTGTACTCT 326

RESULT 9
OY 9
ID Q27354 standard; DNA; 464 BP.
AC Q27354;
DE 03-FEB-1993 (first entry)
DE Encodes A5B7 gH-1 antibody grafted heavy chain.
KM humanised antibody; chimaeric; carcino-embryonic antigen; therapy;
KM diagnosis; carcinomas; CDR: 196; human; murine; ss.
OS Chimaeric.
FH Key Location/Qualifiers
FT cds 22.457
FT /*tag= a
FT /*tag= b
FT /*tag= "murine grafted CDR1"
FT /*tag= "murine grafted CDR2"
FT /*tag= "murine grafted CDR3"
FT /*tag= d
FT /*tag= "murine grafted CDR3"
FT misc_feature 154.184
FT misc_feature 226.280
FT misc_feature 379.406
FT /*tag= d
FT /*tag= "murine grafted CDR3"
FT W09201059-A.
PN 23-JAN-1992.
PF 05-JUL-1991; G01108.
PR 05-JUL-1990; GB-014932.
PR 21-DEC-1990; WO-602017.
PR 05-JUL-1991; WO-G01108.
PA (CELL-) CELLTRECH LTD.
PI Adair JR, Bodmer MW, Mountain A, Owens RJ;
DR WPI: 92-284316/34.
DR P-PSDB: R26151.
FT Humanised antibody molecules - comprising murine and human regions,
FT specific for carcino-embryonic antigen, useful for diagnosis and
FT therapy
PS Example 4; Figure 8; 71pp; English.
CC This sequence encodes a CDR-grafted A5B7 human antibody heavy chain
CC having murine CDRs at amino acids 26-35 (CDR1), 50-65 (CDR2), and
CC 95-102 (CDR3) and additional murine framework residues at 1, 48, 49, 72,
CC 73, 76 and 93. The LAY framework was chosen when making the construct
CC as it shows the highest homology to A5B7.
SQ Sequence 464 BP; 120 A; 108 C; 133 G; 103 T;

Query Match 12.7%; Score 27.6; DB 1; Length 464;
Best Local Similarity 50.8%; Pred. No. 5.2;
Matches 66; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
```

```
QY 36 CATCAGAAATGAGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGATGATTAACC 95
DB 197 CACCTGGAAAGGAGCTCGAGTGGCTTCAATCGAAATAGGCAATGGTACACAA 256
QY 96 TCCAGAACAGCCAGTGGATGATGACAGATGATCGTGCAGACACCGTGGGAAAGACA 155
DB 257 CAGAGTACTCTGCATCTGTGGAAGGGAAGATTCACATTTCCAGAGACAAGACAACTCCA 316
QY 156 CTTTGTTTCT 165
DB 317 CACTGTACT 326

RESULT 10
X10506/c
ID X10506 standard; DNA: 150 BP.
AC X10506;
DE 30-MAR-1999 (first entry)
KW Human biallelic polymorphic DNA fragment WI-13578.
KW Polymorphism: biallelic; human; forensic; paternity testing; disease;
KW detection: phenotypic typing; characteristic: infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
KW treatment; marker; ss.
OS Homo sapiens.
PN M09820165-A2.
PD 14-MAY-1998; U20313.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (MHED) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI: 98-28697/25.
DE New isolated nucleic acid segments from the human genome - used for
DE determining polymorphic forms for use in e.g. forensics, paternity
DE testing or phenotypic typing for disease
PT Claim 1; Page 56; 310pp; English.
PS X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as ataxia, Huntington's disease, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
CC Sequence 150 BP; 49 A; 30 C; 26 G; 44 T;

Query Match 12.7%; Score 27.6; DB 1; Length 150;
Best Local Similarity 60.8%; Pred. No. 3.5;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
```

```
AC T98688;
DE 09-NOV-1998 (first entry)
DE DNA encoding a S. pneumoniae protein of unknown function.
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KW pathogenesis; ss.
OS Streptococcus pneumoniae.
FH Key
FT CDS
FT Location/Qualifiers
FT CDS
FT /*tag= a

PN M09743303-A1.
PD 20-NOV-1997.
PF 14-MAY-1997; U07950.
PR 14-MAY-1996; US-017670.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI Stodola RK;
DR WPI: 98-008793/01.
DR P-PSDB: W38642.
PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections
PS Claim 4; Page 198; 483pp; English.
PS This sequence encodes a Streptococcus pneumoniae protein of unknown
CC function, and represents a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.
CC Sequence 769 BP; 204 A; 202 C; 140 G; 223 T;

Query Match 12.6%; Score 27.4; DB 1; Length 769;
Best Local Similarity 55.9%; Pred. No. 7.3;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 68 ATCCAGCTTGAGTAAGGATTAACCTCCGAAACAGCCAGTGTATGACACAGTG 127
DB 559 ATCTGATAGACTTCAATGCCATTTCCGCAATCTTCAGTAATAGCTTCACTGTA 618
QY 128 ATCTGAGACAGCCGTGGGAAGACACTTG 160
DB 619 ACCATGCTCTGCACCGCCGACAAAGAGAAATTG 651

RESULT 12
X30779
ID X30779 standard; DNA: 769 BP.
AC X30779;
DE 20-MAY-1999 (first entry)
DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:56.
KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;
KW streptococcal infection; pneumococcal; ss.
OS Streptococcus pneumoniae.
PN M09737026-A1.
PD 09-OCT-1997.
PF 01-APR-1997; U05306.
PR 22-AUG-1996; US-025788.
```

PR 02-APR-1996; US-014690.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,  
 PI Stodola RK;  
 DR WPI: 97-503111/46.  
 DR P-PSDB: Y11181.  
 PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
 PT vaccines, drug screening, etc  
 PS Claim 5; Page 91: 354pp; English.  
 CC X30724 to X30946 represent genomic DNA sequences isolated from  
 CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
 CC encode the novel proteins given in Y1114 to Y1137. The proteins,  
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
 CC streptococcal infections and in assays for identifying compounds that  
 CC inhibit or activate the activity of the proteins. The antagonists can  
 CC be used to treat an individual having need to inhibit a bacterial  
 CC protein. Vectors expressing the proteins can be used to induce a  
 CC protective immune response in mammals.  
 SQ Sequence 769 BP; 204 A; 202 C; 140 G; 223 T;

Query Match 12.6%; Score 27.4; DB 1; Length 769;  
 Best Local Similarity 55.9%; Pred. No. 7.3;  
 Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 68 ATCCAGCTTGAGAGTAAGGATTACCTCCAGAACACCGTAGTGAATGACACAGTG 127  
 DB 559 ATCTCGATGAGAGTTCAATGATCCCTCCCAATCTTCAAGTAATAGCTTCACTGA 618  
 OY 128 ATCGTGACAGCACCCTGGGAAAGACACTTG 160  
 DB 619 ACCATGGTCTGCACCCGCCAGACAGAGATTGG 651

## RESULT 13

Q27946 13  
 ID Q27946 standard; cDNA; 787 BP.

AC Q27946;

DT 28-JAN-1993 (first entry)

DE Sequence encoding P6 precursor protein.

KM Monocyte chemoattractant; bovine P6-derivative; thrombosis; tumour;

OS Inflammation therapy; ss.

OS Bos taurus.

FH Key Location/Qualifiers

FT misc\_signal 42..48

FT /tag- a

FT /label- translation consensus sequence

FT poly-a\_signal 746..751

FT signal\_peptide 45..113

FT /tag- b

FT mat\_peptide 114..344

FT /tag- d

FT DE4125251-C.

PD 03-SEP-1992.

PR 31-JUL-1991; 125251.

PR 31-JUL-1991; DE-125251.

PA (SCHM-) SCHAPER & BRUEMME GMBH & CO KG.

PI Gramm W, Lins E;

DR WPI: 92-293438/36.

DR P-PSDB: R26381.

PT Dry containing a bovine protein homologous to human MCP-1 - for

PT treating inflammation, tumours, thrombosis, and immune reactions,

PT also for diagnosis

PS Claim 2; Page 4; 6pp; German.

CC The expression vector lambda gtl1. 1.5 x 10(5) cDNA clones were

CC screened with a polyclonal anti-P6 antiserum of monospecific

CC immunoglobulin G and six positives were identified. The insert of a

CC suitable cDNA clone, p42, was cloned into pUC18 and sequenced.

CC p42 encodes the 11,114 Da precursor of P6. It is called Monocyte

CC Chemoattractant (MCP-1), which is a homologue of human (h)MCP-1.

CC There is 72% overall AA sequence homology to hMCP-1 with the signal  
 CC peptide showing 100% and the central region showing 89% homology.  
 SQ Sequence 787 BP; 236 A; 162 C; 144 G; 245 T;

Query Match 12.3%; Score 26.8; DB 1; Length 787;  
 Best Local Similarity 55.3%; Pred. No. 12;  
 Matches 52; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

OY 60 AGCGCTCCATCCAGCTTGAGAGTAAGGATTACCTCCAGAACACCGTAGTGAATG 119  
 DB 169 AGATCTCCATCCAGAGGCTGATGACTAGACAGAGACACACAGAGTGTCTAAAG 228  
 OY 120 GCACGTGATCTGGACAGCAGCAGCTGGGAAGCA 153  
 DB 229 AGCGTGTGATTTCAGACCATCTCTGGCAAGCA 262

## RESULT 14

Q75094/C  
 ID Q75094 standard; cDNA; 423 BP.

AC Q75094;

DT 11-AUG-1995 (first entry)

DE Plasmid pOKSc4c contg. 3' end of psychosis protecting protein gene.

KM Psychosis protecting protein; PCR; primer: ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT misc\_difference 1..30

FT /tag- a

FT /note- "PCR primer used to generate cDNA library"

FT misc\_difference 402..423

FT /tag- b

FT /note- "see above"

PN WO9426107-A.

PD 24-NOV-1994.

PF 13-MAY-1994; U05445.

PR 13-MAY-1993; US-060560.

PA (UNYNY) UNIV NEW YORK STATE.

PI Basham DA, Friedhoff AJ, Miller JC;

DR WPI: 95-006234/01.

PT New nucleic acids encoding psychosis protecting peptide and

PT antibodies - for the treatment, diagnosis and research of

PT psychotic disorders, such as schizophrenia

PS Claim 1; Page 50-51; 87 pp; English.

CC A subtracted cDNA library was generated from the RNA of lymphocytes

CC obtd. from monozygotic twins discordant for schizophrenia. The two

CC oligos used in library construction were T7 promoter and Sp6

CC promoter (see Q75091, Q75092). A cDNA library was made for each

CC twin and subtractive hybridisation was achieved. Both libraries

CC were used as driver and substrate in two separate subtraction

CC assays. Clone pOKSc was differentially expressed. The expression

CC of the clone was greatest in the 'well' twin. It contains the 3'

CC of a previously unreported gene. It has the closest homology with

CC Varicella-Zoster virus, as 80.0% identity in 26 bp overlap. This

CC cDNA is also expressed in rodent brain in the following areas:

CC cortex, hippocampus and medial geniculate nucleus.

CC Sequence 423 BP; 122 A; 91 C; 102 G; 108 T;

Query Match 12.3%; Score 26.8; DB 1; Length 423;  
 Best Local Similarity 61.4%; Pred. No. 9.2;  
 Matches 43; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 102 ACAGCCAGTGTATTAAGGCACAGTGTGCGACACACCGTGGGAAGACACTTGT 161

DB 217 ACCGCCAGTAGTGAAGACAGTGTCTTCTGTCGCGACCACTAGTGGCTTGTAGG 158

OY 162 TTCTATCAC 171

DB 157 ACCTTATCAC 148

RESULT 15



V34275  
ID V34275 standard; DNA: 1088 BP.  
AC V34275:  
DT 28-JAN-1999 (first entry)  
DE Human secreted protein gene 69 clone HPEBD70.  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;  
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
KW Homo sapiens.  
OS Wo9839446-A2.  
PN 11-SEP-1998.  
PD 06-MAR-1998: U04492.  
PE 07-MAR-1997: US-038621.  
PR 07-MAR-1997: US-040161.  
PR 07-MAR-1997: US-040162.  
PR 07-MAR-1997: US-040163.  
PR 07-MAR-1997: US-040333.  
PR 07-MAR-1997: US-040334.  
PR 07-MAR-1997: US-040336.  
PR 07-MAR-1997: US-040626.  
PR 11-APR-1997: US-043311.  
PR 11-APR-1997: US-043312.  
PR 11-APR-1997: US-043313.  
PR 11-APR-1997: US-043314.  
PR 11-APR-1997: US-043315.  
PR 11-APR-1997: US-043568.  
PR 11-APR-1997: US-043569.  
PR 11-APR-1997: US-043576.  
PR 11-APR-1997: US-043580.  
PR 11-APR-1997: US-043580.  
PR 11-APR-1997: US-043670.  
PR 11-APR-1997: US-043671.  
PR 11-APR-1997: US-043672.  
PR 11-APR-1997: US-043674.  
PR 23-MAY-1997: US-047492.  
PR 23-MAY-1997: US-047500.  
PR 23-MAY-1997: US-047501.  
PR 23-MAY-1997: US-047502.  
PR 23-MAY-1997: US-047503.  
PR 23-MAY-1997: US-047581.  
PR 23-MAY-1997: US-047582.  
PR 23-MAY-1997: US-047583.  
PR 23-MAY-1997: US-047584.  
PR 23-MAY-1997: US-047585.  
PR 23-MAY-1997: US-047586.  
PR 23-MAY-1997: US-047587.  
PR 23-MAY-1997: US-047588.  
PR 23-MAY-1997: US-047589.  
PR 23-MAY-1997: US-047590.  
PR 23-MAY-1997: US-047592.  
PR 23-MAY-1997: US-047593.  
PR 23-MAY-1997: US-047594.  
PR 23-MAY-1997: US-047595.  
PR 23-MAY-1997: US-047596.  
PR 23-MAY-1997: US-047597.  
PR 23-MAY-1997: US-047598.  
PR 23-MAY-1997: US-047599.  
PR 23-MAY-1997: US-047600.  
PR 23-MAY-1997: US-047601.  
PR 23-MAY-1997: US-047612.  
PR 23-MAY-1997: US-047613.  
PR 23-MAY-1997: US-047614.  
PR 23-MAY-1997: US-047615.  
PR 23-MAY-1997: US-047617.  
PR 23-MAY-1997: US-047618.  
PR 23-MAY-1997: US-047632.  
PR 23-MAY-1997: US-047633.

PR 06-JUN-1997: US-048964.  
PR 22-AUG-1997: US-048974.  
PR 22-AUG-1997: US-056630.  
PR 22-AUG-1997: US-056631.  
PR 22-AUG-1997: US-056632.  
PR 22-AUG-1997: US-056636.  
PR 22-AUG-1997: US-056637.  
PR 22-AUG-1997: US-056642.  
PR 22-AUG-1997: US-056644.  
PR 22-AUG-1997: US-056645.  
PR 22-AUG-1997: US-056845.  
PR 22-AUG-1997: US-056862.  
PR 22-AUG-1997: US-056864.  
PR 22-AUG-1997: US-056872.  
PR 22-AUG-1997: US-056874.  
PR 22-AUG-1997: US-056875.  
PR 22-AUG-1997: US-056876.  
PR 22-AUG-1997: US-056877.  
PR 22-AUG-1997: US-056878.  
PR 22-AUG-1997: US-056879.  
PR 22-AUG-1997: US-056880.  
PR 22-AUG-1997: US-056881.  
PR 22-AUG-1997: US-056882.  
PR 22-AUG-1997: US-056884.  
PR 22-AUG-1997: US-056886.  
PR 22-AUG-1997: US-056887.  
PR 22-AUG-1997: US-056888.  
PR 22-AUG-1997: US-056889.  
PR 22-AUG-1997: US-056892.  
PR 22-AUG-1997: US-056893.  
PR 22-AUG-1997: US-056894.  
PR 22-AUG-1997: US-056903.  
PR 22-AUG-1997: US-056908.  
PR 22-AUG-1997: US-056909.  
PR 22-AUG-1997: US-056910.  
PR 22-AUG-1997: US-056911.  
PR 05-SEP-1997: US-057650.  
PR 05-SEP-1997: US-057651.  
PR 05-SEP-1997: US-057651.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,  
Feng P, Ferlie AM, Fischer CL, Graves KA, Greene JM, Hu JS,  
Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,  
Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z,  
WPI: 98-609887/51.  
DR P-PDB: W75178.  
PT New isolated human genes and the secreted polypeptides they encode  
PT - useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
PT Claim 1: Page 277; 447bp: English.  
CC This sequence represents a nucleic acid molecule which encodes a secreted  
CC human protein. The gene number, and the clone it is derived from, are  
CC detailed in the descriptor line. The gene can be used to generate fusion  
CC proteins by linking to the gene to a human immunoglobulin Fc portion  
CC (e.g. V3415) for increasing the stability of the fused protein as  
CC compared to the human protein only.  
CC The invention relates to 70 novel genes and their fragments (nucleic acid  
CC sequences: V34154-V34275; amino acid sequences W75057-W75179) which  
CC are useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. Also, pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the 70 polynucleotides, based on  
CC which tissues they are most highly expressed in (see V34154 for described  
CC uses).  
SQ Sequence 1088 BP; 232 A; 314 C; 294 G; 241 T;

Query Match 12.3%; Score 26.8; DB 1; Length 1088;  
Best Local Similarity 52.7%; Pred. No. 13;  
Matches 58; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
OY 51 CTGCTCTCAGCGCTCCATCCAGCTGAGAGTAAGGATTAACTCCAGACAGCCACT 110  
DB 799 CTGCTCTCAGCGCTCCAGAGAGTAAGGATTAACTCCAGACAGCCACT 858

Oy 111 GGATGATGGACAGTGAATCGTGACAGCACCGTGGAAAGGACACTTTG 160  
| | | | | | | | | | | | | | | | | | | | | |  
Db 859 GTTATCTGGCCCCATGACCGTGGCCACAGCCCTGCTCCAGCAGCACTTG 908

Search completed: August 6, 1999, 04:03:14  
Job time: 1546 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:41 ; Search time 3059.95 Seconds

(without alignments)  
140.529 Million cell updates/sec

Title: us-09-049-696-9

Perfect score: 218  
Sequence: 1 GAACATGCGCCATGATG.....TCTGTGGATCCAGTGGAC 218

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST.\*  
1: em\_est1.\*  
2: em\_est2.\*  
3: em\_est3.\*  
4: em\_est4.\*  
5: em\_est5.\*  
6: em\_est6.\*  
7: em\_est7.\*  
8: em\_est8.\*  
9: em\_est9.\*  
10: em\_est10.\*  
11: em\_est11.\*  
12: em\_est12.\*  
13: em\_est13.\*  
14: em\_est14.\*  
15: em\_est15.\*  
16: em\_est16.\*  
17: em\_est17.\*  
18: em\_est18.\*  
19: em\_est19.\*  
20: em\_est20.\*  
21: em\_est21.\*  
22: em\_est22.\*  
23: em\_est23.\*  
24: em\_est24.\*  
25: em\_est25.\*  
26: em\_est26.\*  
27: em\_est27.\*  
28: em\_est28.\*  
29: em\_est29.\*  
30: em\_est30.\*  
31: em\_est31.\*  
32: em\_est32.\*  
33: em\_est33.\*  
34: em\_est34.\*  
35: em\_est35.\*  
36: em\_est36.\*  
37: em\_est37.\*  
38: em\_est38.\*  
39: em\_est39.\*  
40: em\_est40.\*  
41: em\_est41.\*  
42: em\_est42.\*  
43: em\_est43.\*  
44: em\_est44.\*  
45: em\_est45.\*  
46: em\_est46.\*  
47: em\_est47.\*  
48: em\_est48.\*  
49: em\_est49.\*  
50: em\_est50.\*  
51: em\_est51.\*  
52: em\_est52.\*  
53: em\_est53.\*

54: em\_est22.\*  
55: em\_est23.\*  
56: em\_est24.\*  
57: em\_est25.\*  
58: em\_est26.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	26	11.9	455	39	AA871197	AA871197 Vg32c11.r
2	19	8.7	397	25	W16532	W16532 zb10b12.r1
3	19	8.7	293	31	AA321170	AA321170 EST32647
4	18	8.3	511	31	AA311390	AA311390 EST182307
5	18	8.3	316	34	AA486171	AA486171 ab14f02.r
6	18	8.3	404	36	AA659068	AA659068 nu81d02.s
7	17	7.8	416	20	D21547	D21547 M0568H09.mo
8	17	7.8	313	20	T52594	T52594 yb21c03.r1
9	17	7.8	101	21	T61187	T61187 yb45c09.r1
10	17	7.8	589	27	AA003372	AA003372 mc47g09.r
11	17	7.8	365	27	W91192	W91192 mt72d07.r1
12	17	7.8	437	28	AA064183	AA064183 mj64f03.r
13	17	7.8	384	29	AA125957	AA125957 zn27e08.r
14	17	7.8	213	29	AA137598	AA137598 mq86f12.r
15	17	7.8	453	29	AA177303	AA177303 mt22f11.r
16	17	7.8	275	32	AA344252	AA344252 EST50133
17	17	7.8	565	34	AA464650	AA464650 zx85c03.r
18	17	7.8	545	35	AA538001	AA538001 vj34g07.r
19	17	7.8	511	35	AA543238	AA543238 vk37g10.r
20	17	7.8	428	41	AI006744	AI006744 ua82h12.r
21	17	7.8	333	43	AI163725	AI163725 A047P14U
22	17	7.8	641	43	AI166346	AI166346 xylem.est
23	17	7.8	634	43	AI166438	AI166438 xylem.est
24	17	7.8	423	43	AI174114	AI174114 vz87e08.r
25	17	7.8	444	43	AI225702	AI225702 uf10e08.y
26	17	7.8	132	44	AI183330	AI183330 ta42d07.x
27	17	7.8	467	45	AI328465	AI328465 aa403ne.f
28	17	7.8	576	48	AI591843	AI591843 mt22f11.y
29	16	7.3	461	21	R06611	R06611 yf10f06.s1
30	16	7.3	488	22	R52606	R52606 y81h02.r1
31	16	7.3	460	22	R64455	R64455 y136h02.r1
32	16	7.3	423	22	R72171	R72171 y188d06.s1
33	16	7.3	286	22	R74940	R74940 M0B0304R.MO
34	16	7.3	379	22	R78053	R78053 y180b04.r1
35	16	7.3	409	23	H18587	H18587 yn51b09.r1
36	16	7.3	449	23	H18868	H18868 yn52a10.r1
37	16	7.3	458	23	H19977	H19977 yn53h09.r1
38	16	7.3	501	23	R90959	R90959 yp92e04.r1
39	16	7.3	493	23	R90963	R90963 yp92f03.r1
40	16	7.3	301	24	D19163	D19163 M0SG00204
41	16	7.3	470	24	H83354	H83354 yv82f03.r1
42	16	7.3	450	24	H86398	H86398 ys93b07.r1
43	16	7.3	550	24	N35748	N35748 yx81c04.r1
44	16	7.3	450	25	N93279	N93279 zb68g08.s1
45	16	7.3	478	51	AI735151	AI735151 as68e12.x

#### ALIGNMENTS

RESULT 1  
AA871197  
LOCUS  
DEFINITION  
IMAGE:1095956 5' similar to SW:ECIC\_BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.  
ACCESSION  
AA871197  
NID  
92966642

AA871197 455 bp mRNA EST 16-MAR-1998  
Vg32c11.r1 Barstead bowel MRLR9 Mus musculus cDNA clone

VERSION AA871197.1 GI:2966642  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 455)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMT Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1395394.

CONTACT: Marra M/Mouse EST Project  
WashU-HMT Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:602188  
Seq primer: -28ml3 rev2 ET from Amer sham  
High quality sequence stop: 414.

FEATURES  
source  
Location/Qualifiers  
1..455  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1095956"  
/clone\_lib="Barstead bowel MPLRB9"  
/tissue\_type="bowel"  
/dev\_stage="8 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TCTTACGATCTGACGTGAGCGAGCGCCGCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[AATCGATCTG], digested with Not I and cloned into the  
Not I and Eco RI sites of the modified pT73 vector.  
Source irradiated bowel harvested 72 hours after  
irradiation (1400 Gys). Library constructed by Bob  
Barstead."

BASE COUNT 122 a 123 c 105 g 105 t  
ORIGIN

Query Match 11.9%; Score 26; DB 39; Length 455;  
Best Local Similarity 100.0%; Pred. No. 0.00016;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 TTGTTTCTTATCAGCTGACACGCA 183  
|||||  
Db 161 TTGTTTCTTATCAGCTGACACGCA 186

RESULT 2  
W16532 397 bp mRNA EST 29-APR-1996  
LOCUS zbl0812.r1 Soares\_fetal\_lung\_NbHL19W Homo sapiens cDNA clone  
DEFINITION IMAGE:301631 5', mRNA sequence.  
ACCESSION W16532  
NID 91290914  
VERSION W16532.1 GI:1290914  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 397)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On May 8, 1995 this sequence version replaced gi:800920.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: ETPrimer  
High quality sequence stop: 317.

FEATURES  
source  
Location/Qualifiers  
1..397  
/organism="Homo sapiens"  
/db\_xref="GDB:1246555"  
/db\_xref="taxon:9606"  
/clone="IMAGE:301631"  
/clone\_lib="Soares\_fetal\_lung\_NbHL19W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTACCAATCTGACGTGAGCGAGCGCCGCAATTTTCTTTTCTTTT  
3']; double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NbHL19W."

BASE COUNT 109 a 95 c 110 g 82 t 1 others  
ORIGIN

Query Match 8.7%; Score 19; DB 25; Length 397;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 AATCCTTCTCTGGATCCC 211  
|||||  
Db 386 AATCCTTCTCTGGATCCC 368

RESULT 3  
AA321170/c 293 bp mRNA EST 19-APR-1997  
LOCUS AA321170 Frontal lobe Homo sapiens cDNA 5' end, mRNA sequence.  
DEFINITION AA321170  
ACCESSION AA321170  
NID 91973518  
VERSION AA321170.1 GI:1973518  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 293)  
Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.  
Rapid cDNA sequencing (expressed sequence tags) from a  
directionally cloned human infant brain cDNA library  
Nature Genet. 4, 373-380 (1993)  
JOURNAL MEDLINE  
94004965

COMMENT On Sep 12, 1996 this sequence version replaced gi:1398058.

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423

Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse

## FEATURES

## Source

1. 293  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):121740"  
/db\_xref="taxon:9606"  
/clone\_lib="Frontal lobe"  
/dev\_stage="adult"  
/note="Organ: brain; Vector: pbluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"

BASE COUNT 103 a 42 c 88 g 58 t 2 others  
ORIGIN

Query Match 8.7%; Score 19; DB 31; Length 293;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 CAGCCTCCCAATCCTTC 200  
|||||  
Db 154 CAGCCTCCCAATCCTTC 136

RESULT 4  
AA311390 511 bp mRNA EST 19-APR-1997  
LOCUS EST182307 Jurkat T-cells VI Homo sapiens CDNA 5' end, mRNA  
DEFINITION  
sequence.  
AA311390  
ACCESSION  
91963904  
NID  
VERSION  
AA311390.1 GI:19663904  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 511)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
On Sep 12, 1996 this sequence version replaced gi:1400909.  
Other ESTs: THC184526  
Contact: Kerlavage, AR

Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423

Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

## FEATURES

## Source

1. 511  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):158802"  
/db\_xref="taxon:9606"  
/clone\_lib="Jurkat T-cells VI"  
/cell\_type="T-lymphocyte"  
/note="Vector: pbluescript SK-; Site\_1:  
XhoI"

BASE COUNT 124 a 136 c 157 g 92 t 2 others  
ORIGIN

Query Match 8.3%; Score 18; DB 31; Length 511;  
Best Local Similarity 100.0%; Pred. No. 5.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 194 ATCCTCTCTGGATCCC 211  
|||||  
Db 410 ATCCTCTCTGGATCCC 393

RESULT 5  
AA486171 316 bp mRNA EST 06-MAR-1998  
LOCUS ab14f02.r1 Striatogene lung (#937210) Homo sapiens CDNA clone  
DEFINITION  
IMAGE:840795 5', mRNA sequence.  
AA486171  
ACCESSION  
92216387  
NID  
VERSION  
AA486171.1 GI:2216387  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 316)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Insert Length: 889 Std Error: 0.00  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 200.  
location/Qualifiers

## FEATURES

## Source

1. 316  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="814H06; 14"  
/clone="IMAGE:840795"  
/clone\_lib="Striatogene lung (#937210)"  
/sex="male"

/dev\_stage="72 years"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: lung; Vector: pBluescript SK-; Site: 1:  
 EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:  
 oligo dt. normal lung. Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GATTCGCGACAG  
 3' -3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGT 3' "

BASE COUNT 84 a 81 c 103 g 48 t

ORIGIN

Query Match 8.3%; Score 18; DB 34; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 194 ATCCTCTCTGGGATCCC 211  
 ||||||||||||||||  
 DB 150 ATCCTCTCTGGGATCCC 133

RESULT 6  
 AA659068 404 bp mRNA EST 03-DEC-1997  
 LOCUS nu81d02.s1 NCI-CGAP\_Alvi1 Homo sapiens cDNA clone IMAGE:1217091,  
 DEFINITION mRNA sequence.  
 ACCESSION AA659068  
 NID 92595222  
 VERSION AA659068.1 GI:2595222  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 404)  
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On May 5, 1995 this sequence version replaced gi:798066.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D.  
 cDNA Library Preparation: David B. Krizman, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/dbfp/image/image.html

Insert Length: 949 Std Error: 0.00  
 Seg primer: -40m13 fwd. ET from AmerSham  
 High quality sequence stop: 396.

## FEATURES

Source

Location/Qualifiers

1..404

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="11"

/clone\_lib="NCI-CGAP\_Alvi1"

/tissue\_type="alveolar rhabdomyosarcoma"

/lab\_host="DH10B"

/note="Vector: PAMP10; mRNA made from alveolar  
 rhabdomyosarcoma, cDNA made by oligo-dT priming.  
 Non-directionally cloned. Size-selected on agarose gel,  
 average insert size 600 bp. Reference: Krizman et al.  
 (1996) Cancer Research 56:5380-5383."

BASE COUNT 104 a 104 c 140 g 56 t

ORIGIN

Query Match 8.3%; Score 18; DB 36; Length 404;  
 Best Local Similarity 100.0%; Pred. No. 5.6;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 194 ATCCTCTCTGGGATCCC 211  
 ||||||||||||||||  
 DB 154 ATCCTCTCTGGGATCCC 137

RESULT 7  
 D21547 416 bp mRNA EST 07-OCT-1996  
 LOCUS M0568H09 mouse embryonal carcinoma cell line F9 Mus musculus cDNA  
 DEFINITION clone 68H09, mRNA sequence.

ACCESSION D21547  
 NID 9618684  
 VERSION D21547.1 GI:618684  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 416)  
 REFERENCE Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T. and Shimada, K.  
 A survey of genes expressed in undifferentiated mouse embryonal  
 carcinoma F9 cells: characterization of low-abundance mRNAs  
 J. Biochem. 116, 128-139 (1994)  
 95096008

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

Contact: Kazunori Shimada  
 Department of Medical Genetics, Division of Molecular Biomedicine  
 Research Institute for Microbial Diseases, Osaka University  
 3-1, Yamadaoka, Suita, Osaka, 565, Japan  
 Tel: 06-879-8325  
 Fax: 06-879-8326.

FEATURES  
 source Location/Qualifiers  
 1..416  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="68H09"  
 /clone\_lib="mouse embryonal carcinoma cell line F9"

BASE COUNT 87 a 96 c 116 g 117 t

ORIGIN

Query Match 7.8%; Score 17; DB 20; Length 416;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 CCTTCTCTGGGATCCC 212  
 ||||||||||||||||  
 DB 353 CCTTCTCTGGGATCCC 337

RESULT 8  
 T52594 313 bp mRNA EST 06-FEB-1995  
 LOCUS yb2ic03.r1 Striatogene fetal spleen (#937205) Homo sapiens cDNA  
 DEFINITION clone IMAGE:71812 5', mRNA sequence.

ACCESSION T52594  
 NID 9654454  
 VERSION T52594.1 GI:654454  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 313)  
 REFERENCE Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, B.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,

Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
High quality sequence stops: 170  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: M13RP1  
High quality sequence stop: 170.  
Location/Qualifiers  
1. 313  
/organism="Homo sapiens"  
/db\_xref="GDB:493477"  
/db\_xref="taxon:9606"  
/map="22"  
/clone="IMAGE:71812"  
/clone\_lib="Stratagene fetal spleen (#937205)"  
/tissue\_type="fetal spleen"  
/dev\_stage="fetal"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="organ: spleen; Vector: pBluescript SK-; Site,1:  
EcoRI; Site,2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Hepatectomy from normal male caucasian. Average  
insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' CTCGAGTGTGTGTGTGTGTGT 3'."  
3' -3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGT 3'."  
2 others

BASE COUNT 45 a 86 c 77 g 103 t  
ORIGIN

Query Match 7.8%; Score 17; DB 20; Length 313;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 TCCTTTGGGCGCCCTT 35  
|||||  
Db 249 TCCTTTGGGCGCCCTT 265

RESULT 9  
T61187/c 101 bp mRNA EST 13-FEB-1995  
LOCUS yca5c09.r1 Stratagene liver (#937224) Homo sapiens cDNA clone  
IMAGE:83632 5' similar to contains Alu repetitive element; mRNA  
sequence.  
T61187  
ACCESSION T61187  
NID 9664224  
VERSION T61187.1 GI:664224  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 101)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,  
Chaisson, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohtling, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 70  
High quality sequence stops: 83 Source: IMAGE Consortium, LNL This  
clone is available royalty-free through LNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.  
Insert length: 70 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 83.  
Location/Qualifiers  
1. 101  
/organism="Homo sapiens"  
/db\_xref="GDB:500689"  
/db\_xref="taxon:9606"  
/clone="IMAGE:83632"  
/clone\_lib="Stratagene liver (#937224)"  
/sex="male"  
/dev\_stage="49 years old"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="organ: liver; Vector: pBluescript SK; Site,1:  
EcoRI; Site,2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Hepatectomy from normal male caucasian. Average  
insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GAATCGGCGACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTGTGTGTGTGTGT 3'."

BASE COUNT 26 a 24 c 28 g 23 t  
ORIGIN

Query Match 7.8%; Score 17; DB 21; Length 101;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 CAGCTCCCAATCCT 198  
|||||  
Db 40 CAGCTCCCAATCCT 24

RESULT 10  
AA003372 589 bp mRNA EST 19-JUL-1996  
LOCUS mg47g09.r1 Soares mouse embryo NM013.5 14.5 Mus musculus cDNA  
DEFINITION clone IMAGE:426976 5' similar to gb:U17502 Mouse mRNA overexpressed  
and amplified in teratocarcinoma cell (MOUSE); mRNA sequence.  
AA003372  
ACCESSION AA003372  
NID 91446837  
VERSION AA003372.1 GI:1446837  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 589)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theisinger, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On May 8, 1995 this sequence version replaced gi:799522.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:261528

Seq primer: ETP-Primer  
High quality sequence stop: 312.  
Location/Qualifiers

# FEATURES

source

```
1. .589
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:426976"
/clone_1lb="Soares mouse embryo NBM13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGGAGCGCGCGGAAATTTTATTTTATTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
```

BASE COUNT 102 a 134 c 149 g 204 t

ORIGIN

Query Match 7.8%; Score 17; DB 27; Length 589;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 AGGACACTTGTCTTCTT 166  
|||||  
Db 96 AGGACACTTGTCTTCTT 112

RESULT 11  
W91192/c 365 bp mRNA EST 12-SEP-1996  
LOCUS m72d07.r1 Soares mouse embryo NBM13.5 14.5 Mus musculus cDNA  
DEFINITION clone IMAGE:419821 5', mRNA sequence.  
W91192  
ACCESSION G1538899  
NID W91192.1 GI:1538899  
VERSION W91192.1 GI:1538899  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 365)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Trelsing,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407026.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.

MGI:254373  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 353.  
Location/Qualifiers

# FEATURES

source

```
1. .365
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:419821"
/clone_1lb="Soares mouse embryo NBM13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGGAGCGCGCGGAAATTTTATTTTATTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
```

BASE COUNT 89 a 100 c 89 g 87 t

ORIGIN

Query Match 7.8%; Score 17; DB 27; Length 365;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 AGGATTACCCCTCAG 100  
|||||  
Db 221 AGGATTACCCCTCAG 205

RESULT 12  
AA064183 437 bp mRNA EST 24-SEP-1996  
LOCUS m64f03.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone  
DEFINITION IMAGE:480893 5', mRNA sequence.  
AA064183  
ACCESSION G1558174  
NID AA064183.1 GI:1558174  
VERSION AA064183.1 GI:1558174  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 437)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Trelsing,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On May 5, 1995 this sequence version replaced gi:798450.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:291637  
Seq primer: -28M13 rev2 from Amersham



High quality sequence stop: 428.

## FEATURES

source

Location/Qualifiers

1. 437

/organism="Mus musculus"

/db\_xref="taxon:10090"

/map="21"

/clone="IMAGE:480893"

/clone\_lib="Soares mouse p3NMF19.5"

/dev\_stage="19.5 dpc total fetus"

/lab\_host="DH10B (ampicillin resistant)"

/note="Vector: pT73D (Pharmacia) with a modified polylinker. Site 1: Not I. Site 2: Eco RI. 1st strand cDNA was primed with a Not I oligo (GT primer [5' TGTTACCAATCTGAGTGGGAGCGCGCATTTTCTTTTCTTTT 3']

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

## BASE COUNT

102 a 129 c 123 g 83 t

## ORIGIN

Query Match 7.8%; Score 17; DB 28; Length 437;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 AACCTCCAGACAGCC 107

Db 73 AACCTCCAGACAGCC 89

## RESULT 13

AA125957

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

AA125957 384 bp mRNA EST 23-DEC-1997  
 zn27e08.r1 Stratagene neuroepithelium NT2RAM1 937234 Homo sapiens  
 cDNA clone IMAGE:548678 5', mRNA sequence.  
 AA125957 91685623  
 AA125957.1 GI:1685623  
 EST.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 384)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Chissone, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevisan, E., Underwood, K., Woldmann, P., Waterston, R., Wilson, R.  
 and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 On Sep 12, 1996 this sequence version replaced gi:1297522.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

WARNING: There is evidence that suggests that the 384-well parent  
 plate of this clone contains both human and mouse derived clones.  
 Thus, the origin of this clone is uncertain. This caution should be  
 kept in mind should you use this clone.

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 791 Std Error: 0.00  
 Seq primer: -28M13 rev2 from AmerSham  
 High quality sequence stop: 70.  
 Location/Qualifiers

## FEATURES

source

1. 384

/organism="Homo sapiens"

/db\_xref="GDB:3927194"

/db\_xref="taxon:9606"

/clone="IMAGE:548678"

/clone\_lib="Stratagene neuroepithelium NT2RAM1 937234"

/dev\_stage="Ntera-2/RA+MI neuroepithelial cells"

/lab\_host="SOLR (kanamycin resistant)"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
 XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2  
 (Ntera-2/cl.D1) precursor cells induced with Retinoic  
 Acid for 1 week, followed by 3 weeks in mitotic inhibitors  
 (Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR  
 vector; -5' adaptor sequence: 5' GAATTCGACAG 3' -3'  
 adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

## BASE COUNT

90 a 76 c 90 g 126 t 2 others

## ORIGIN

Query Match 7.8%; Score 17; DB 29; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 GAAATGAGCTGTCTT 58

Db 229 GAAATGAGCTGTCTT 245

## RESULT 14

AA137598

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

AA137598 213 bp mRNA EST 09-FEB-1997  
 mg86f12.r1 Stratagene mouse melanoma (#937312) Mus musculus CDNA  
 clone IMAGE:585647 5', mRNA sequence.  
 AA137598 91699636  
 AA137598.1 GI:1699636  
 EST.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 213)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 On Sep 12, 1996 this sequence version replaced gi:1393890.

Contact: Maria M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.

MGI:360295

Seq primer: -28M13 rev1 ET from AmerSham

High quality sequence stop: 212.

## FEATURES

source

1. 213

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:585647"

/clone\_lib="Stratagene mouse melanoma (#937312)"

```

/tissue_type="melanoma"
/dev_stage="M2 cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1;
EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:
Oligo dT. From M2 cells, a highly metastatic derivative of
the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGACGACG
3' ~3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGT 3'."

```

```

BASE COUNT      52 a      55 c      52 g      54 t
ORIGIN
Query Match      7.8%; Score 17; DB 29; Length 213;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 AGGATTACCTCCAG 100
|||||
Db 149 AGGATTACCTCCAG 133

```

## RESULT 15

```

AA177303      453 bp      mRNA      EST      16-FEB-1997
LOCUS      mt22f11.r1 Soares mouse 3NDXS Mus musculus cDNA clone IMAGE:621837
DEFINITION      5' similar to TR:G563127 G563127 ACID FINGER PROTEIN. ; mRNA
sequence.
ACCESSION      AA177303
NID      G1758604
VERSION      AA177303.1 GI:1758604
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE      1 (bases 1 to 453)
AUTHORS      Maria M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

```

```

TITLE      The WashU-HMT Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1393242.

```

```

Contact: Maria M/Mouse EST Project
WashU-HMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through INL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:382661
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 424.
Location/Qualifiers
1. 453

```

## FEATURES

```

source
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:621837"
/clone_lib="Soares mouse 3NDXS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pTR3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAGTGGAGCGCGCGCTTTTGTGTGTGTGTGTGTGTGTGTGT

```

```

3'}. double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTR3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT      109 a      146 c      105 g      93 t
ORIGIN
Query Match      7.8%; Score 17; DB 29; Length 453;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 196 CCTCTCTGGATCCCA 212
|||||
Db 306 CCTCTCTGGATCCCA 322

```

```

Search completed: August 6, 1999, 09:22:44
Job time: 6235 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:07 ; Search time 3075.15 seconds  
(without alignments)  
146.890 Million cell updates/sec

Title: US-09-049-696-10

Perfect score: 229

Sequence: 1 GGCACAGTGTGCTGACAG.....AACCTGACCTGACTGTCA 229

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	161.8	70.7	455	39	AA871197	AA871197 vq32c11.r
2	63	27.5	479	35	AA541829	AA541829 vj01f01.r
3	34.2	14.9	373	33	AA443218	AA443218 aa14d10.s
4	34.2	14.9	435	33	AA443258	AA443258 aa14d09.r
5	34.2	14.9	467	40	AA932194	AA932194 om84f02.s
6	34.2	14.9	430	47	AI493356	AI493356 tg70f11.x
7	34	14.8	496	22	R44318	R44318 yq35b09.s1
8	32.4	14.1	492	24	N31560	N31560 yx59h03.r1
9	32.2	14.0	386	21	T78952	T78952 yd21a08.r1
10	32	13.9	760	41	AI053077	AI053077 BSMMP520
11	31.8	13.8	468	37	AA726662	AA726662 vu93f05.r
12	31.6	13.8	753	30	AA208962	AA208962 mo80e02.r
13	29.8	13.0	667	45	AI346181	AI346181 qp49b03.x
14	29.6	12.9	438	21	R07193	R07193 yf14f01.s1
15	29.6	12.9	309	21	R07194	R07194 yf14f02.s1
16	29.6	12.9	124	22	R25684	R25684 yq44d09.r2
17	29.6	12.9	667	34	AA519171	AA519171 TGEESTz34
18	29.6	12.9	555	34	AA519694	AA519694 TGEESTz27
19	29.6	12.9	616	34	AA520304	AA520304 TGEESTz38
20	29.6	12.9	573	34	AA531884	AA531884 TGEESTz32
21	29.6	12.9	427	46	AA926371	AA926371 UI-R-AI-e
22	29.4	12.8	441	40	AA944727	AA944727 EST200226
23	29.2	12.8	496	22	H07987	H07987 y186g02.r1
24	29.2	12.8	502	22	R67110	R67110 y131c06.r1
25	29.2	12.8	390	30	AA230091	AA230091 nc49f12.s
26	29.2	12.8	447	34	AA529995	AA529995 vj14g04.r
27	29.2	12.8	502	39	AA861972	AA861972 o145e03.s
28	29.2	12.7	400	23	H30430	H30430 ym58a12.r1
29	29	12.7	451	27	W82768	W82768 tF03b08.r1
30	29	12.7	412	28	AA068843	AA068843 mm39g11.r
31	29	12.7	604	28	AA097854	AA097854 mk11c09.r
32	29	12.7	483	33	AA389037	AA389037 mp16c10.r
33	29	12.7	577	35	AA537195	AA537195 vk45f02.r
34	29	12.7	492	35	AA590057	AA590057 v057c07.r
35	29	12.7	486	36	AA611120	AA611120 v061f01.r
36	29	12.7	500	38	AA755827	AA755827 vu04c06.r
37	29	12.7	428	38	AA790583	AA790583 vx71g08.r
38	29	12.7	378	39	AA869799	AA869799 vq16a01.r
39	29	12.7	469	39	AA893375	AA893375 EST197178
40	29	12.7	441	42	AI158557	AI158557 ud25f11.r
41	29	12.7	413	45	AI325676	AI325676 mm39g11.y
42	29	12.7	680	47	AI527737	AI527737 uJ28e09.y
43	28.8	12.6	601	27	AA037278	AA037278 zc52c04.x
44	28.8	12.6	453	31	AA291957	AA291957 zt45h03.s
45	28.8	12.6	397	39	AA820138	AA820138 LD23605.5

ALIGNMENTS

RESULT 1  
LOCUS AA871197 455 bp mRNA  
DEFINITION vq32c11.r1 Barstead bowel MLR9 Mus musculus cDNA clone  
IMAGE:1095956 5' similar to SW:ECIC\_BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.  
ACCESSION AA871197  
NID g2966642

VERSION	AA871197.1	GI:2966642
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
REFERENCE	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 455) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R.	
TITLE	The MashU-HM Mouse EST Project	
JOURNAL	Unpublished (1996)	
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1395394.	
	Contact: Marra M/Mouse EST Project MashU-HM Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LINDA; contact the IMAGE Consortium (info@image.lindl.gov) for further information. MGI:602188 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 414. Location/Qualifiers 1..455 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone_1="IMAGE:1095956" /clone_1ib="Barstead bowel MPLRB9" /tissue_type="bowel" /dev_stage="8 weeks" /lab_host="DH10B" /note="vector: pRT7D-Pac (pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGATCTGAAAGTGGAGGCGGCCCTTTTCTTTTCTTTTCTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors [AATTCGATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Source irradiated bowel harvested 72 hours after irradiation (1400 Gys). Library constructed by Bob Barstead."	
BASE COUNT	122 a 123 c 105 g 105 t	
ORIGIN		
Query Match	70.7%; Score 161.8; DB 39; Length 455;	
Best Local Similarity	81.7%; Pred. No. 4.7e-43;	
Matches 187; Conservative	0; Mismatches 42; Indels 0; Gaps 0;	
QY	1 GGCACAGTGATCGTGACAGACACCGTGGAAAGACACTTTGTTCTATACCTGGACA 60	
DB	122 GGCATAGTGATCGTGACAGACCTCGGGGCAAGACACCTTGTCTTACCTGGACA 181	
QY	61 AGCGAGCCGCCCAATCCTCTCTGGGGATCCCAAGTGGACAGAGCAAGTGGCTTGT 120	
DB	182 AGCGATTCCTCTACATATTATTCTGGGATCCCAAGCGGAGTGGACAAATGGTTTATA 241	
QY	121 GTGGACAAAACCAAAATGGCTACCTCCAAATCCCAAGGATTCCTAAGTTGGCACT 180	
DB	242 CTAGACACACCACTAAGTGGGCTACTCTCAAGTCCCAAGGACGCGCTAAGGTGGCTT 301	
QY	181 TGGAAATACAGTCTGCAGACAGCTCACAACTTACCTTACTACTCA 229	
DB	302 TGGAAATACAGCATTCACGAGGACTACAGACTCTTACCTTGACTTCA 350	

RESULT	2
LOCUS	AAS41829
DEFINITION	vj01f01.1 Barstead mouse pooled organs MRLRB4 Mus musculus cDNA clone IMAGE:920473 similar to TR.GI184066 GI184066 CACIM-ACTIVATED CHLORIDE CHANNEL. ; mRNA sequence.
ACCESSION	AAS41829
NID	q2288263
VERSION	AAS41829..1 GI:2288263
KEYWORDS	EST.
SOURCE	Mus house. house muscul Mus musculus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 479) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubucque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,W., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The MashU-HMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1406841.  Contact: Marra M/Mouse EST Project MashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LINL : contact the IMAGE Consortium ( <a href="#">infoimage.linl.gov</a> ) for further information. GSI:532688 Seq primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 368.  location/qualifiers 1..479 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:920473" /clone_lib="Barstead mouse pooled organs MRLRB4" /seq_type="mixed" /_tissue_type="pooled organs" /_dev_stage="7 day" /_lab_host="DH10B" /_note="Organ: pooled; Vector: pT7mD-Pac (Pharmacia) with a modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTAGGGCGGCCCCCTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTGGATTTCGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7m3 vector. Library constructed by Bob Barstead."
FEATURES	
SOURCE	
BASE COUNT	139 a 115 c 120 g 105 t
ORIGIN	
Query Match	27.5%; Score 63; DB 35; Length 479;
Best Local Similarity	82.8%; Pred. No. 1.4e+10;
Matches	Conservative 72; Mismatches 15; Indels 0; Gaps 0;
OY	143 CCTACTCCTCAAAATCCCAGGCATGCTAAGGTGGCACCTTGGAAATAACAGCTCGACAAGCAA 202
Dd	 5 CCACCTCCCAAGTCCCAAGCAGCGTAAGGTGCTTTGGAAAATACAGCATTTACGAGA 64
OY	203 GC TCACA AAC CT TG AC CG T GA CT GT CA 229
Dd	 65 GC TC ACA AG A CT CTC ACC TT G ACT GT CA 91

```

LOCUS      AA443218/c                               EST                               03-JUN-1997
DEFINITION aa44d10.sl Soares_NbHMPu_S1 Homo sapiens cDNA clone IMAGE:813235 3'
            similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.
ACCESSION  AA443218
NID        92155893
VERSION    AA443218.1 GI:2155893
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE  1 (bases 1 to 373)
            Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisler,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,D.,
            Moore,B., Schellenberg,K., Stepien,M., Tan,F., Theising,B.,
            White,Y., Wyllie,T., Waterston,R. and Wilson,R.
            WashU-Merck EST Project 1997
            Unpublished (1997)
TITLE      JOURNAL
COMMENT    Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -41m13 fwd. ER from Amersham
            High quality sequence stop: 359.
FEATURES   Location/Qualifiers
            1..373
             /organism="Homo sapiens"
             /db_xref="GDB:6044145"
             /db_xref="taxon:9606"
             /map="788H12; 14q24.3"
             /clone="IMAGE:813235"
             /clone_id="Soares_NbHMPu_S1"
             /tissue_type="Soares_NbHMPu_S1"
             /tissue_type="Pooled human melanocyte, fetal heart, and
             pregnant uterus"
             /lab_host="DH10B"
             /note="Organ: mixed (see below); Vector: pT73D-Pac
             (Pharmacia) with a modified polylinker; Site_1: Not I;
             Site_2: Eco RI; Equal amounts of plasmid DNA from three
             normalized libraries (melanocyte 2NbHw, pregnant uterus
             NbHwU, and fetal heart NbH19W) were mixed, and ss circles
             were made in vitro. Following HAP purification, this DNA
             was used as tracer in a subtractive hybridization
             reaction. The driver was PCR-amplified cDNAs from pools of
             5,000 clones made from the same 3 libraries. The pools
             consisted of I.M.A.G.E. clones 260232-265223,
             340488-345479, and 484488-489479."
BASE COUNT 117 a 71 c 71 g 114 t
ORIGIN
Query Match 14.9%; Score 34.2; DB 33; Length 373;
Best Local Similarity 76.4%; Pred.No.0.36;
Matches 42; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Oy 3 CACACTGATCGTGAGCAGCAGCCGCTGGGAAGACACTTGTTCATTACCTCGG 57
    ||||| |||| | ||||| ||||| ||||| ||||| |||||
Db 314 CACACTGACTGTGATTAATAGTCTGGGCACACACTATGTTCTAGTACTCGG 260
RESULT 4
LOCUS      AA443258 435 bp mRNA EST 03-JUN-1997
DEFINITION aa14d09.r1 Soares_NbHMPu_S1 Homo sapiens cDNA clone IMAGE:813233 5'
            similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.
            ; mRNA sequence.

```

ACCESSION	AA443258
NUM	92155933
VERSION	AA443258.1
KEYWORDS	GI:21555933
SOURCE	EST.
ORGANISM	human.
REFERENCE	Human sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 435)
JOURNAL	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maita, M., Martin, J., Moore, B., Schellenberg, K., Septon, M., Tan, F., Teising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
COMMENT	Mashu-Merck EST Project 1997 Unpublished (1997)
FEATURES	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -28ml3 rev2 E9 from Amersham High quality sequence stop: 412. Location/Qualifiers 1..435 /organism="Homo sapiens" /db_xref="GDB:6044143" /db_xref="taxon:9606" /clone="IMAGE:813233" /clone_1db="Soares_NHMFPU.S1" /tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus" /lab_host="DH10B" /note="Organ: mixed (see below); Vector: pT7TD-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBH, pregnant uterus NBH9, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."
BASE COUNT	131 a 86 c 86 g 132 t
ORIGIN	
Query Match	14.9%; Score 34.2; DB 33; Length 435;
Best Local Similarity	76.4%; Pred. No. 0.38;
Matches	42; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY	3 CACACTGATCGTGGACAGCCGTGGGAAGACACTTGTTCCTAACCTCG 57  Db 201 CACACTGACTGTGATTAATAGTGTGGCAGACACTATGTTCTAGTACGTGG 255
RESULT	5
LOCUS	AA932194 467 bp mRNA EST 17-JUN-1998
DEFINITION	om84f02.s1 NCI CGAP Kid3 Homo sapiens cDNA clone IMAGE:1553883 3'
ACCESSION	AA932194
NUM	93087106
VERSION	AA932194.1
KEYWORDS	GI:3087106
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.















**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:40:43 ; Search time 650.92 Seconds

(without alignments)  
88.020 Million cell updates/sec

Title: US-09-049-696-10

Perfect score: 229  
Sequence: 1 GGACAGTGCATGCTGGACAG.....AACCTTGACCTGACTGCA 229

Scoring table: OLIGO\_MWC

Searched: 311585 segs, 125096042 residues

Database : N.Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	7.9	2040	T43949	Sequence flanking
2	17	7.4	843	Q10677	Modified hepatitis
3	17	7.4	276	T26704	Human gene signatu
4	17	7.4	325	V78785	Staphylococcus aur
5	17	7.4	305	V78656	Staphylococcus aur
6	17	7.4	14078	V74502	Staphylococcus aur
7	17	7.4	357	V78411	Staphylococcus aur
8	16	7.0	5187	T14218	Mouse patched gene
9	16	7.0	2237	T38267	200 gene different
10	16	7.0	5187	V21589	Mouse patched (plc
11	16	7.0	5187	V21587	Precis coenia (but
12	16	7.0	3046	V29074	Rattus norvegicus
13	16	7.0	5187	V64092	Mouse patched gene
14	16	7.0	5187	V64099	Precis coenia patc
15	16	7.0	529	V87937	EST clone FG708. N
16	16	7.0	438	X41096	Human secreted pro
17	15	6.6	134525	004525	Total base sequenc
18	15	6.6	134525	004525	Total base sequenc
19	15	6.6	775	M40220	DNA sequence inclu
20	15	6.6	5270	Q25387	Rat thyrotropin re
21	15	6.6	2401	Q48539	Bacillus subtilis
22	15	6.6	1633	Q51227	Human MSH-R gene.
23	15	6.6	1270	044340	Sequence encoding
24	15	6.6	378	080358	HCT-15 contg. a DN
25	15	6.6	1428	Q85873	B70 type B antigen
26	15	6.6	36333	088003	Ad2-ORF6/PGK-CFTR
27	15	6.6	1166	T04604	CYP19 gene 1.4 pro
28	15	6.6	505	T04607	CYP19 gene exon 1.
29	15	6.6	195	T01072	Human B7-2 exon h1
30	15	6.6	305	T12166	Partial pDGA-5-CDK
31	15	6.6	2101	T31451	Drosophila Ndr ser
32	15	6.6	11000	T58840.1	Continuation (2 of
33	15	6.6	633	T67391	H. pylori inner me
34	15	6.6	717	T67884	H. pylori inner me
35	15	6.6	951	T68792	Melanocortin-1-rec
36	15	6.6	2803	T62384	Human A33 antigen
37	15	6.6	2565	T62369	Human A33 antigen
38	15	6.6	2044	T75623	Canine herpesvirus
39	15	6.6	44377	T78508	Platenolide syntha
40	15	6.6	44377	T80414	Platenolide syntha
41	15	6.6	951	V06397	Human melanocortin
42	15	6.6	1633	V19136	Human melanocyte s
43	15	6.6	1448	T84017	DNA encoding a sul

# ALIGNMENTS

RESULT 1	
T43949	15 6.6 729 1 V22744
ID T43949 standard; DNA; 2040 BP.	
AC T43949:	
DT 18-AUG-1997 (first entry)	
DE Sequence flanking marker 63-2 in HH region of chromosome 6p2.1.	
KW Primer; polymerase chain reaction; amplification; hereditary haemochromatosis;	
KW HH; mutation; HH-associated allele; base-pair polymorphism; HHP-1;	
KW HHP-19; HHP-29; microsatellite repeat allele; genetic marker;	
KW interferon treatment; hepatitis C infection; ss.	
OS Synthetic.	
PN W09635803.AL.	
PD 14-NOV-1996.	
PE 08-MAY-1996; U06583.	
PF 08-MAY-1995; US-436074.	
PR 15-NOV-1995; US-559302.	
PO 09-FEB-1996; US-599252.	
PA (MERC-) MERCATOR GENETICS INC.	
PI Drayna DT, Feder JN, Gairke A, Kimmel BE, Thomas WJ;	
PI Wolff RK;	
DR WPI: 96-518691/51.	
PT Diagnosing and genotyping of hereditary haemochromatosis (HH) -	
PT using primers to detect specific polymorphisms of the HH gene on	
PT chromosome 6p2.1 or novel microsatellite markers	
PS Dislosure; Fig 1Y; 67pp; English.	
CC The sequences given in T43925-55 represent portions of the genome	
CC surrounding several markers of the invention. The markers were	
CC identified using the series of primer pairs given in T71901-72	
CC which were used to determine the presence or absence of the common	
CC hereditary haemochromatosis (HH) gene mutation in an individual. The	
CC method comprised assessing genomic DNA from an individual for the	
CC presence or absence of the HH-associated allele of the single base-pair	
CC polymorphism HHP-1, HHP-19 or HHP-29, and/or at least one non-optional	
CC marker comprising the following microsatellite repeat alleles of group	
CC A and optionally of group B:	
CC Group A: 19D9, 18B4, 1A2, 1E4, 24E2, 2B8, 321-1, 4073-1, 4440-1, 4440-2,	
CC 731-1, 5091-1, 3216-1, 4072-2, 950-1, 950-2, 950-3, 950-5, 950-6,	
CC 950-8, 63-1, 63-2, 63-3, 65-1, 65-2, 373-8, 373-29, 68-1, 241-6, 241-29;	
CC Group B: D6S464, D6S306, D6S258, D6S265, D6S105 and D6S1001.	
CC The absence of the genotype indicates the likelihood of the presence of	
CC the HH mutation. Knowledge of the new genetic markers allows the	
CC definition of genotypes characteristic of heterozygous carriers and	
CC homozygotes having a HH mutation in their genomic DNA. The potential for	
CC HH in an individual interferes with the effectiveness of interferon	
CC treatment for hepatitis C infection. By diagnosing this potential, the	
CC responsiveness of interferon treatment may be evaluated.	
SO Sequence 2040 BP; 713 A; 409 C; 398 G; 513 T;	
Query Match	7.9%; Score 18; DB 1; Length 2040;
Best Local Similarity	100.0%; Pred. No. 1.7;
Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 114 CTTTGAGTGCACAAA 131	
DB 195 CTTTGAGTGCACAAA 212	
RESULT 2	
ID Q10677/c	
AC Q10677:	
DT 29-APR-1991 (first entry)	
DE Modified hepatitis B virus large protein (III).	
KW Hepatitis B virus; large surface protein; L protein; myristylation;	
KW vaccines; ss.	

OS Hepatitis B virus.  
 FH Key location/Qualifiers  
 FT cds 1. 123.  
 FT /\*tag- a  
 FT /label= pre-S1\_region  
 FT cds 124. 162  
 FT /\*tag- b  
 FT /label= pre-S2\_region  
 FT cds 163. 843  
 FT /\*tag- C  
 FT /label= S-protein  
 FT  
 FT EP-414374-A.  
 FT PD 27-FEB-1991.  
 FT PE 19-JUL-1990: 307900.  
 FT PR 25-JUL-1989: US-385342.  
 FT PR 03-AUG-1989: US-389184.  
 FT PA (SMIRK) SMITHKLINE BIOLOGIC.  
 FT PI Comberbach M, Harford N, Cabezon T, Rutgers A, Voet P,  
 PI Jacobs E, Hollenberg CP, Janowicz ZA, Merckelberg AJ.  
 PI WPI: 91-059585/09.  
 DR P-PSDB: R10850.  
 DR Modified hepatitis B virus large surface protein - has amino acid  
 PT sequence encoding L protein used in vaccine for treating or  
 PT preventing hepatitis B without effects  
 PS Disclosure: Page 4-7: 79pp: English.  
 CC The modified protein comprises amino acids 12-52 (amino acid 13 is  
 CC opt. deleted), followed by 133-145, then 175-400 of the wild-type  
 CC sequence.  
 CC This protein is devoid of the following properties: to function  
 CC as a substrate to yeast glycosylation enzymes, polymerised human  
 CC serum albumin binding capacity and protease sensitivity.  
 CC The modified product may be used alone or in a composite particle  
 CC comprising at least two polypeptides corresp. to all or part of a  
 CC protein having the biological activity of one of the hepatitis  
 CC B surface antigens. The composite particles are useful for  
 CC preparing improved hepatitis B vaccines or for treating of HBV.  
 CC See also Q10677-79.  
 SO Sequence 843 BP: 174 A: 229 C: 179 G: 261 T:

Query Match 7.4%; Score 17; DB 1; Length 843;  
 Best Local Similarity 100.0%; Pred. No. 5.6;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 CTCTGGGATCCAGTGG 98  
 ||||||||||||||||  
 DB 133 CTCTGGGATCCAGTGG 117

RESULT 3  
 T26704  
 ID T26704 standard; cDNA to mRNA; 276 BP.  
 AC T26704:  
 DT 16-OCT-1996 (first entry)  
 DE Human gene signature HMG508954.  
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 OS Homo sapiens.  
 PN MO9514772-A1.  
 PD 01-JUN-1995.  
 PF 11-NOV-1994: J01916.  
 PR 12-NOV-1993: JP-355504.  
 PA (MATS/) MATSUBARA K.  
 PA (OKUB/) OKUBO K.  
 PI Matsubara K, Okubo K;  
 PI WPI: 95-206931/27.  
 DR Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 PS Claim 1: Page 2152-2153: 2245pp: Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in T19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SO Sequence 276 BP: 65 A: 77 C: 63 G: 69 T:

Query Match 7.4%; Score 17; DB 1; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 GGACAGAGCAAGGTGG 113  
 ||||||||||||||||  
 DB 187 GGACAGAGCAAGGTGG 203

RESULT 4  
 ID V78785 standard; DNA; 325 BP.  
 AC V78785:  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus confg SEQ ID #4474.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PF 07-JAN-1997: 100117.  
 PR 05-JAN-1996: US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunach CA,  
 PI Rosen CA;  
 PI WPI: 97-374922/35.  
 DR Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1: Page 2979: 3271pp: English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 SO Sequence 325 BP: 103 A: 55 C: 62 G: 102 T:

Query Match 7.4%; Score 17; DB 1; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 TTGTACTGACAAAAC 132  
 DB 171 TTGTAGTGACAAAAC 187

RESULT 5  
 V78656/c  
 ID V78656 standard; DNA: 305 BP.  
 AC V78656;  
 DE 16-MAR-1999 (first entry)  
 KM Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KM skin infection; surgical wound infection; scalded skin syndrome;  
 KM toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PF 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 PI WPI; 97-374922/35.  
 DR Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1; Page 2927; 3271BP; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 SQ Sequence 305 BP; 92 A; 48 C; 48 G; 114 T;

Query Match 7.4%; Score 17; DB 1; Length 305;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 TTGTACTGACAAAAC 132  
 DB 55 TTGTAGTGACAAAAC 39

RESULT 6  
 V74502/c  
 ID V74502 standard; DNA: 14078 BP.  
 AC V74502;  
 DE 16-MAR-1999 (first entry)  
 KM Staphylococcus aureus contig SRO ID #191.  
 DE Staphylococcus aureus contig SRO ID #191.  
 KM Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KM skin infection; surgical wound infection; scalded skin syndrome;  
 KM toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 FH Key location/Qualifiers  
 FT misc-feature 1381..1440  
 FT /tag-a

FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"

FT misc-feature 3181..3240  
 FT /tag-b  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"

FT misc-feature 4981..5040  
 FT /tag-c  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"

FT misc-feature 6781..6840  
 FT /tag-d  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"

FT misc-feature 8581..8640  
 FT /tag-e  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"

FT misc-feature 10381..10440  
 FT /tag-f  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"

FT misc-feature 12181..12240  
 FT /tag-g  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"

FT misc-feature 13981..14040  
 FT /tag-h  
 FT /note- "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"

PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PF 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 PI WPI; 97-374922/35.  
 DR Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1; Page 887-895; 3271BP; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences

CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.  
SQ Sequence 14078 BP; 4438 A; 2440 C; 2355 G; 4363 T;

Query Match 7.4%; Score 17; DB 1; Length 14078;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 TTGTAGTGACAAAAC 132  
Db 6739 TTGTAGTGACAAAAC 6723

RESULT 7  
V78411/c  
ID V78411 standard; DNA: 357 BP.

AC 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #4100.  
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
OS Staphylococcus aureus.  
PN EP-786519-A2.  
PD 30-JUL-1997.  
PE 07-JAN-1997; 100117.  
PR 05-JAN-1996; US-009861.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
PI Rosen CA;  
PI WPI: 97-374922/35.

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
PS Claim 1: Page 2625; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S.aureus DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against S.aureus infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.  
CC Sequence 357 BP; 103 A; 69 C; 59 G; 123 T;

Query Match 7.4%; Score 17; DB 1; Length 357;  
Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 TTGTAGTGACAAAAC 132  
Db 215 TTGTAGTGACAAAAC 199

RESULT 8  
T14218/c  
ID T14218 standard; cDNA: 5187 BP.  
AC T14218;  
DN 11-JUL-1996 (first entry)

DE Mouse patched gene.  
KW Patched gene; ptc gene; embryo development; cellular regulation;  
KW signal transduction; ligand; antibody; hedgehog protein;  
KW gene therapy; ss.  
OS Mus musculus.

FN Key cds Location/Qualifiers  
FT 97..4401  
FT /\*tag- a  
FT misc\_difference 4536..4537  
FT /\*tag- b  
FT /note- "bases 4536 and 4537 are given as k in the  
FT specification"

FT misc\_difference 4563  
FT /\*tag- c  
FT /note- "base 4563 is given as r in the  
FT specification"

PN W09611260-A1.  
PD 18-APR-1996.  
PP 06-OCT-1995; U13233.  
PR 07-OCT-1994; US-319745.  
PI (STRD ) UNIV LEHMAN STANFORD JUNIOR.  
PI Goodrich LV, Johnson RL, Scott MP;  
PI WPI: 96-209842/21.  
DR P-PSDB: R94380.

PT DNA encoding patched protein other than Drosophila patched protein -  
PT used to produce antibodies which detect or inhibit patched protein  
PT ligand signal transduction in cells  
PS Disclosure: Page 43-46; 70pp; English.  
CC A cDNA clone (T14218) coding for the mouse patched protein (PTC)  
CC (R94380) was obtd. by PCR amplification of cDNA derived from  
CC a mouse limb bud, and use of the PCR product to probe a mouse 8.5  
CC dpc lambda-gt10 cDNA library. The PCR primers (see T14226-27)  
CC were based on patched amino acid sequences from fly, mosquito (see  
CC also R94384), butterfly (see also R75373) and beetle (see also  
CC R94385). The mouse ptc gene can be used for prodn. of large amounts  
CC of recombinant PTC, as a probe e.g. to detect gene mutations, in  
CC gene therapy, to study embryo development, to produce transgenic  
CC animal models, etc.  
SQ Sequence 5187 BP; 1162 A; 1411 C; 1332 G; 1279 T;

Query Match 7.0%; Score 16; DB 1; Length 5187;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 CAAAACACCAAAATG 141  
Db 5069 CAAAACACCAAAATG 5054

RESULT 9

ID T38267 standard; cDNA: 2237 BP.  
DB T38267;

DE 29-DEC-1996 (first entry)  
DE 200 gene differentially expressed in T helper cells.  
KW T helper cell; TH cell; T-cell; T-lymphocyte; 200 gene;  
KW differential expression; immune disorder; multiple sclerosis;  
KW asthma; lepromatous leprosy; diagnosis; therapy; receptor; ss.  
OS Homo sapiens.

FN Key cds Location/Qualifiers  
FT 42..947  
FT /\*tag- a

FT signal\_peptide 42..101  
FT /\*tag- b

FT mat\_peptide 102..944  
FT /\*tag- c

FT misc\_difference 1298  
FT /\*tag- d

FT /note- "base 1298 is given as r in the  
FT specification"

PN W09627603-A1.  
PD 12-SEP-1996.



```
PF 01-MAR-1996; U02798.
PR 03-MAR-1995; US-398633.
PR 07-JUN-1995; US-487748.
PA (MILL-) MILLENNIUM PHARM INC.
PI Levinson DA;
DR WPI: 96-433404/43.
P-PSDB: W01049.
PR Genes and their products differentially expressed in T helper cells
PT - useful in diagnosis and treatment of immune disorders, e.g.
PS multiple sclerosis, asthma, lepromatous leprosy, etc.
PS Claim 1; Fig 24; 218pp; English.
CC Novel human gene 200 (T38267) was isolated from a lambda gt11
CC human lymphocyte cDNA library using the murine 200 gene (see also
CC T38265) as probe. The gene codes for a novel cell surface
CC receptor (W01049) of the Ig superfamily class. The 200 gene is
CC expressed at levels many-fold higher in T helper TH1 cell
CC subpopulations than in TH2 cell subpopulations. The gene and its
CC product can be used therapeutically to ameliorate immune disorders
CC and in the diagnosis of such disorders. It can also be used to
CC produce soluble 200 gene-Ig fusions and to breed transgenic animals
CC useful as models of TH cell subpopulation-related disorders.
SQ Sequence 2237 BP; 587 A; 518 C; 525 G; 606 T;
```

```
Query Match 7.0%; Score 16; DB 1; Length 2237;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 149 TCCAAATCCAGCAT 164
    |||||
DB 376 TCCAAATCCAGCAT 391
```

## RESULT 10

```
V21589/c
ID V21589 standard; DNA: 5187 BP.
AC V21589;
DT 25-JUN-1998 (first entry)
DE Mouse patched (ptc) protein encoding cDNA.
KW patched protein; ptc; cancer; tumour suppressor; cell adhesion promoter;
KW wound healing; ageing; mouse; ss.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 97..4401
FT FT /tag= a
FT FT /product= "mouse patched protein"
PN W09745541-A2.
PD 04-DEC-1997.
PE 02-JUN-1997; U09553.
PR 31-MAY-1996; US-656055.
PA (REGC ) UNIV CALIFORNIA.
PA (STRD ) UNIV IELAND STANFORD JUNIOR.
PI Epstein E, Goodrich LV, Johnson RL, Oro A, Scott MP;
DR WPI: 98-032648/03.
P-PSDB: W52199.
PR Patched protein other than Drosophila melanogaster patched protein -
PT used for characterising the phenotype of a tumour
PS Claim 5; Pages 63-66; 86pp; English.
CC This genomic DNA encodes a mouse patched (ptc) protein. This can be used
CC to construct an expression cassette comprising an altered patch or
CC hedgehog gene. The expression cassette comprises a nucleic acid encoding
CC a patched protein other than a Drosophila melanogaster patched protein,
CC or fragment of at least 12 nucleotides in length, as other than an intact
CC chromosome under transcriptional control of a transcriptional initiation
CC region, and a transcriptional termination region, both functional in an
CC expression host. A genetically engineered mammalian cell comprising this
CC expression cassette as an extrachromosomal element or integrated into the
CC genome of the cell can be predisposed to develop basal cell carcinoma as
CC a result of the transfection. By analysing DNA, functional analysis of
CC patched protein function, or by detecting antibody binding to abnormal
CC patched protein, a genetic predisposition to developmental abnormalities
CC and cancer can be diagnosed. This analysis can also be used for
CC characterising the phenotype of a tumour, particularly a carcinoma,
```

```
CC especially a basal cell carcinoma. The methods can also be used for
CC characterising transitional cell carcinoma of the bladder, meningiomas
CC medulloblastomas, etc. The modified cells comprising the expression
CC cassette can be used to determine the role of different exons of the
CC patched gene in oncogenesis, signal transduction, etc. Transgenic animal
CC models created from these cells can be used as animal models for
CC carcinomas of the skin. The patched protein of mosquito, butterfly or
CC beetle or alternatively, a mammalian patched protein of human or mouse
CC can be used to identify ligands or substrates that bind to, modulate, or
CC mimic the action of patched gene. These agents could be used as tumour
CC suppressors, cell adhesion promoters (e.g. in wound healing and ageing).
SQ Sequence 5187 BP; 1165 A; 1412 C; 1338 G; 1270 T;
```

```
Query Match 7.0%; Score 16; DB 1; Length 5187;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 126 CAAAACCCAAATG 141
    |||||
DB 5069 CAAAACCCAAATG 5054
```

## RESULT 11

```
V21587/c
ID V21587 standard; DNA: 5187 BP.
AC V21587;
DT 25-JUN-1998 (first entry)
DE Precis coenia (butterfly) patched (ptc) protein encoding DNA.
KW patched protein; ptc; cancer; tumour suppressor; cell adhesion promoter;
KW wound healing; ageing; Precis coenia; butterfly; ss.
OS Precis coenia.
FH Key Location/Qualifiers
FT CDS 04-DEC-1997.
PD 02-JUN-1997; U09553.
PR 31-MAY-1996; US-656055.
PA (REGC ) UNIV CALIFORNIA.
PA (STRD ) UNIV IELAND STANFORD JUNIOR.
PI Epstein E, Goodrich LV, Johnson RL, Oro A, Scott MP;
DR WPI: 98-032648/03.
PR Patched protein other than Drosophila melanogaster patched protein -
PT used for characterising the phenotype of a tumour
PS Claim 2; Pages 49-51; 86pp; English.
CC This DNA is stated to encode a Precis coenia patched (ptc) protein and
CC is identical to the mouse ptc cDNA. This can be used to construct an
CC expression cassette comprising an altered patch or hedgehog gene. The
CC expression cassette comprises a nucleic acid encoding a patched protein
CC other than a Drosophila melanogaster patched protein, or fragment of at
CC least 12 nucleotides in length, as other than an intact chromosome under
CC transcriptional control of a transcriptional initiation region, and a
CC transcriptional termination region, both functional in an expression
CC host. A genetically engineered mammalian cell comprising this expression
CC cassette as an extrachromosomal element or integrated into the genome of
CC the cell can be predisposed to develop basal cell carcinoma as a result
CC of the transfection. By analysing DNA, functional analysis of patched
CC protein function, or by detecting antibody binding to abnormal patched
CC protein, a genetic predisposition to developmental abnormalities and
CC cancer can be diagnosed. This analysis can also be used for
CC characterising the phenotype of a tumour, particularly a carcinoma,
CC especially a basal cell carcinoma. The methods can also be used for
CC characterising transitional cell carcinoma of the bladder, meningiomas
CC medulloblastomas, etc. The modified cells comprising the expression
CC cassette can be used to determine the role of different exons of the
CC patched gene in oncogenesis, signal transduction, etc. Transgenic animal
CC models created from these cells can be used as animal models for
CC carcinomas of the skin. The patched protein of mosquito, butterfly or
CC beetle or alternatively, a mammalian patched protein of human or mouse
CC can be used to identify ligands or substrates that bind to, modulate, or
CC mimic the action of patched gene. These agents could be used as tumour
CC suppressors, cell adhesion promoters (e.g. in wound healing and ageing).
SQ Sequence 5187 BP; 1165 A; 1412 C; 1337 G; 1270 T;
```

Query Match 7.0%; Score 16; DB 1; Length 5187;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 CAAAACACCAAAATG 141  
|||||  
DB 5069 CAAAACACCAAAATG 5054

RESULT 12  
V29074/c  
ID V29074 standard; DNA; 3046 BP.  
AC V29074;  
DT 11-SEP-1998 (first entry)  
DE Rattus norvegicus NRL gene 5' region.  
KW 5' region; NRL gene; neu-related lipocalin; promoter; therapy;  
KM study; breast cancer; mammary gland; ds.  
OS Rattus norvegicus.  
FH Key Location/Qualifiers  
FT promoter 1154..2967  
FT misc\_feature /tag- a 2967  
FT /note- "putative transcription start site" 2939..2945  
FT TATA\_signal /tag- c 2939..2945  
FT W09815634-A1.  
PD 16-APR-1998.  
PE 05-MAY-1997; U07504.  
PR 07-OCT-1996; US-726725.  
PA (WISC.) WISCONSIN ALUMNI RES FOUND.  
PI Chen K, Gould MN.  
DR WPI; 98-240820/21.  
PT New isolated mammary gland specific promoters - can be used to  
PT promote gene expression in the mammary gland in a constant manner  
PT throughout the estrous cycle  
PS Disclosure; Fig 2; 48pp; English.  
CC The sequence is that of the 5' region of the rat NRL  
CC (neu-related lipocalin) gene. The sequence contains a mammary  
CC gland-specific promoter which can be used for targeting foreign gene  
CC expression to the mammary gland. It can be used in therapy and study  
CC of e.g. breast cancer genes.  
SQ Sequence 3046 BP; 727 A; 723 C; 857 G; 739 T;

Query Match 7.0%; Score 16; DB 1; Length 3046;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 96 TGCACAGACGAGGT 111  
|||||  
DB 2108 TGCACAGACGAGGT 2093

RESULT 13  
V64092/c  
ID V64092 standard; cDNA; 5187 BP.  
AC V64092;  
DT 26-JAN-1999 (first entry)  
DE Mouse patched gene.  
KW Mouse; patched gene; diagnosis; treatment; developmental disorder;  
KM cancer; healing; injured tissue; ptc; spina bifida; Wnt-1 oncogene;  
OS Mus sp.  
FH Key Location/Qualifiers  
FT CDS 97..4401  
FT /tag- a  
FT US5837538-A.  
PD 17-NOV-1998.  
PE 06-OCT-1995; 540406.  
PR 07-OCT-1994; US-319745.  
PA (STRD.) UNIV LELAND STANFORD JUNIOR.

PI Goodrich LV, Johnson RL, Scott MP;  
DR WPI; 99-023461/02.  
DE P-PDS; W72968.  
PT Nucleic acid encoding vertebrate patched protein and related  
PT transformants - used to express poly:peptide(s), useful for  
PT diagnosis and treatment of developmental disorders or cancer, and in  
PT healing of injured tissue  
PS Claim 1; Column 43-48; 38pp; English.  
CC The present sequence represents the mouse patched (ptc) gene. Cells  
CC containing and expressing the ptc gene are used for the recombinant  
CC production of the protein. These in turn are useful: (1) for generating  
CC antibodies (Ab); and (11) to screen for specific-binding ligands  
CC (potential therapeutic agonists and antagonists). The ptc gene, or its  
CC fragments, are used to isolate related sequences from other mammals; to  
CC identify mutations (particularly those associated with genetic diseases  
CC such as spina bifida and other developmental disorders); to monitor  
CC expression levels in testis (to determine relationship with sperm  
CC production) and to isolate 5'-non-coding sequences (used to study  
CC embryonic development and to provide regulated expression of proteins).  
CC The complete gene can be used in gene therapy, including expression of  
CC antisense molecules, and to generate transgenic animals for studies of  
CC embryonic development. Ab are used diagnostically to determine the  
CC ptc protein on cell surfaces and as competitive inhibitors of signal  
CC transduction through the ptc ligand. Cells that have been engineered to  
CC express the ptc protein can be used to promote regrowth and healing of  
CC damaged tissue (e.g. growth of new teeth) and regulation of the ptc  
CC protein expression may be useful in cancer treatment (it may control the  
CC Wnt-1 oncogene).  
SQ Sequence 5187 BP; 1165 A; 1412 C; 1337 G; 1270 T;

Query Match 7.0%; Score 16; DB 1; Length 5187;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 CAAAACACCAAAATG 141  
|||||  
DB 5069 CAAAACACCAAAATG 5054

RESULT 14  
V64099/c  
ID V64099 standard; DNA; 5187 BP.  
AC V64099;  
DT 26-JAN-1999 (first entry)  
DE Precs coena patched gene.  
KW Patched gene; ptc; diagnosis; treatment; developmental disorder;  
KM cancer; healing; injured tissue; spina bifida; Wnt-1 oncogene;  
OS Precs coena.  
FH Key Location/Qualifiers  
FT CDS 97..4401  
FT /tag- a  
FT US5837538-A.  
PD 17-NOV-1998.  
PE 06-OCT-1995; 540406.  
PR 07-OCT-1994; US-319745.  
PA (STRD.) UNIV LELAND STANFORD JUNIOR.  
PI Goodrich LV, Johnson RL, Scott MP;  
DR WPI; 99-023461/02.  
PT Nucleic acid encoding vertebrate patched protein and related  
PT transformants - used to express poly:peptide(s), useful for  
PT diagnosis and treatment of developmental disorders or cancer, and in  
PT healing of injured tissue  
PS Example; Column 21-26; 38pp; English.  
CC The present invention describes vertebrate and invertebrate patched (ptc)  
CC genes. Cells containing and expressing the ptc gene are used for the  
CC recombinant production of the protein. These in turn are useful: (1) for  
CC generating antibodies (Ab); and (11) to screen for specific-binding  
CC ligands (potential therapeutic agonists and antagonists). The ptc gene,  
CC or its fragments, are used to isolate related sequences from other  
CC mammals; to identify mutations (particularly those associated with  
CC genetic diseases such as spina bifida and other developmental disorders);  
CC to monitor expression levels in testis (to determine relationship with  
CC sperm production) and to isolate 5'-non-coding sequences (used to study

Search completed: August 5, 1999, 17:40:45  
 JOD time: 6280 sec

CC embryonic development and to provide regulated expression of proteins).  
 CC The complete gene can be used in gene therapy, including expression of  
 CC antisense molecules, and to generate transgenic animals for studies of  
 CC embryonic development. Ab are used diagnostically to determine the  
 CC ptc protein on cell surfaces and as competitive inhibitors of signal  
 CC transduction through the ptc ligand. Cells that have been engineered to  
 CC express the ptc protein can be used to promote regrowth and healing of  
 CC damaged tissue (e.g. growth of new teeth) and regulation of the ptc  
 CC protein expression may be useful in cancer treatment (it may control the  
 CC Mnt-1 oncogene). The present sequence represents Precis coena  
 CC (butterfly) patched gene, from the present invention.  
 SQ Sequence 5187 BP; 1165 A; 1412 C; 1337 G; 1270 T;

Query Match 7.0%; Score 16; DB 1; Length 5187;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CAAAACACCAAAATG 141  
 ||||||||||||||||  
 DB 5069 CAAAACACCAAAATG 5054

## RESULT 15

VF7937  
 ID VF7937 standard; cDNA: 529 BP.  
 AC VF7937;  
 DT 12-FEB-1999 (first entry)  
 DE EST clone FG708.  
 KW Expressed sequence tag; secreted protein; hematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 OS Homo sapiens.  
 PN WO9845437-A2.  
 PD 15-OCT-1998.  
 PF 10-APR-1998; U06956.  
 PR 10-APR-1997; US-837312.  
 PA (GENY ) GENETICS INST INC.  
 PI Agostino MJ, Jacobs K, Lavallee ER, McCoy JM, Merberg D,  
 PI Racine JA, Spaulding V, Treacy M;  
 DR WPI: 99-070078/06.  
 PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries  
 PS Claim 1; Page 229; 641pp; English.  
 CC The present sequence represents an expressed sequence tag (EST), and is  
 CC a polynucleotide of the invention. The polynucleotides of the invention  
 CC are all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.  
 SQ Sequence 529 BP; 127 A; 137 C; 133 G; 132 T;

Query Match 7.0%; Score 16; DB 1; Length 529;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 TCCAATCCAGCAT 164  
 ||||||||||||||||  
 DB 405 TCCAATCCAGCAT 420

---

This Page Blank (uspto)

Result No.	Score	Query Match	Length	DB	ID	Description
1	31.6	13.8	4771	1	V02186	Human type C lectin
2	29.6	12.9	1438	1	N81177	Insert from pLIFNM
3	28.2	12.3	2832	1	O50631	Ovine factor-1, IS
4	28	12.2	29544	1	T12460	Cytomegalovirus g
5	27.8	12.1	3083	1	V59060	Human ribonucleoti
6	27.8	12.1	468	1	X40616	Human secreted pro
7	27.6	12.1	4213	1	Q90019	Human betaglycan
8	27	11.8	5438	1	Q29263	Human calcium cha
9	27	11.8	2889	1	O51080	Plasmodium falciparum
10	27	11.8	1029	1	X07291	Pig-tailed macaque
11	26.8	11.7	376	1	V7386	Staphylococcus au
12	26.6	11.6	1920	1	Q13830	MSP gene in pUCRS
13	26.6	11.6	1742	1	N00004	Sequence of Influe
14	26.6	11.6	8888	1	X02665	EP-892047 Seq ID
15	26.4	11.5	1963	1	Q13981	Xylinol dehydrogen
16	26.4	11.5	826	1	V28683	Ripening banana pu
17	26.4	11.5	1751	1	V59610	Human secreted pro
18	26.4	11.5	2377	1	V59758	Human secreted pro
19	26.2	11.4	32768	1	X13060	Enterococcus faec
20	26	11.4	1077	1	N91628	Sequence of gamma
21	26	11.4	3664	1	O62675	Plasmodium falciparum
22	26	11.4	3267	1	O88049	Glucan Lyase 1 ge
23	26	11.4	3267	1	O87603	Fungus-infected G
24	26	11.4	777	1	T34680	Penicillin-resista
25	26	11.4	996	1	T36906	FKC-II genes. Pri
26	26	11.4	920	1	T66144	Fv.3418 insert en
27	26	11.4	999	1	T66145	Fv.4715-myc inser
28	26	11.4	920	1	T94607	HindIII-EcoRI ins
29	26	11.4	999	1	T94608	HindIII-EcoRI ins
30	26	11.4	237326	1	V57903	Hereditary haemoch
31	26	11.4	235033	1	V57926	Hereditary haemoch
32	26	11.4	3267	1	X84192	Gracilariaopsis le
33	26	11.4	878	1	X30341	DNA encoding a hu
34	25.8	11.3	346	1	O60383	Human brain expres
35	25.8	11.3	10366	1	O67401	Huntingtin DNA
36	25.8	11.3	10366	1	T97924	Composite sequenc
37	25.8	11.3	10366	1	V05828	Human huntingtin
38	25.8	11.3	32768	1	X13037	Enterococcus faec
39	25.8	11.3	634	1	X30403	DNA encoding a hu
40	25.6	11.2	3336	1	O35992	Human pemphigus v
41	25.6	11.2	1088	1	T71217	Type I, p80 IL-1-1
42	25.6	11.2	535	1	T83860	DNA encoding a Sta
43	25.6	11.2	3159	1	V43962	Seq ID #1 from DE

Seq ID #2 from DE1  
streptococcus pneu

## ALIGNMENTS

RESULT	1
ID	V02186
AC	V02186; standard; cDNA; 4771 BP.
DT	12-MAY-1998 (first entry)
DE	Human type C lectin encoding cDNA.
KW	Human; type C lectin; E-selectin; competitive inhibitor;
KW	molecular marker; ss.
OS	Homo sapiens.
FT	Key
FT	location/Qualifiers
FT	12..4452
FT	/*tag= a
FT	/product= "type C lectin"
PN	M09740154-A1.
PD	30-OCT-1997.
PE	17-APR-1997; U06347.
PR	24-APR-1996; U5-637021.
PA	(GETH ) GENENTECH INC.
P1	Laaky LA; Wu K;
DR	WPI: 97-535838/49.
DR	P-PSDB: W44119.
PT	Human and mouse type C lectin(s) - useful as competitive inhibitor
PT	of lectin activity and as molecular markers for tissues that express
PT	them
PS	Example 1; Page 48-51; 97pp; English.
CC	The present sequence encodes a novel type C lectin. Type C lectins
CC	can be used to identify and purify their native ligands and compete
CC	with them for their binding, especially useful as competitive
CC	inhibitors of the biological activity of native type C lectins.
CC	This makes type C lectins useful as molecular markers for tissues
CC	in which they are expressed. The nucleic acid, encoding type C lectin
CC	can be used to identify other type C lectin coding sequences.
SO	Sequence 4771 BP; 944 A; 1517 C; 1498 G; 812 T;

Query Match	13.8%;	Score 31.6;	DB 1;	Length 4771;
Best Local Similarity	53.2%;	Pred. No. 0.3;		
Matches 67;	Conservative	0;	Mismatches 59;	Indels 0;
			Gaps	0.

[illegible]

RESULT	2	
ID	N81177/c	
ID	N81177	standard; CDNA; 1438 BP.
AC	N81177;	
DT	06-NOV-1990	(first entry)
DE	Insert from pLIFN3 encoding leukemia inhibitory factor. LIF; ss	
KW	Plasmod; pLIFN3; pLIF7.2b; leukaemia inhibitory factor.	
OS	Mus musculus.	
FH	Key	Location/Qualifiers
FT	signal_peptide	23..94
FT		/*tag= a
FT		95..631
FT	mat_peptide	/*tag= b
FT		/product=LIF
FT		

PN EP-285448-A.  
 PD 05-OCT-1988.  
 PF 31-MAR-1988; 302962.  
 PR 02-APR-1987; AU-001209.  
 PR 24-JUL-1987; AU-003317.  
 PR 15-OCT-1987; AU-004903.  
 PR 21-DEC-1987; AU-006005.  
 PA (AMRA) Amrad Corp Ltd.  
 PI Gearing DP, Gough NM, Hilton DJ, King JA, Metcalf D, Nice EC,  
 PI Nicola NA, Simpson RJ, Willson TA;  
 DR WPI: 88-280202/40.  
 DR P-PSDB: P80577.  
 PT Leukemia-inhibitory factor - used for suppressing myeloid leukaemia  
 PT and modifying macrophage function and other responses to infections.  
 PS Example 6; Fig 15; 53pp; English.  
 CC Clones pLIF7.2b and pLIFNK1 (N81170 and N81171 resp.) contain incomplete  
 CC copies of murine LIF coding sequence. Neither contains the complete  
 CC hydrophobic leader sequence. In order to isolate a clone contg. the  
 CC entire coding sequence, two probes were used. One corresponded to  
 CC nucleotides 67-102 at the 3' end of pLIF7.2b; the second corresponded to  
 CC nucleotides 500-535 at the 5' end. Clone lambdaDLIFNK3, which hybridised  
 CC with both probes, was subcloned into pEMBL8+ to generate pLIFNK3.  
 CC See also N81169-79.  
 SO Sequence 1438 BP; 343 A; 366 C; 383 G; 344 T; 2 Others;

Query Match 12.9%; Score 29.6; DB 1; Length 1438;  
 Best Local Similarity 61.8%; Pred. No. 0.91;  
 Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 59 CAAAGCAGCCCTCCCAATCTCTCTGAGATCCAGTGCAGACAGAGAGTGGCTTG 118  
 DB 1098 CAAACCAAGCCTTCAACATTTCTCTGAGACAGACAGCAAGACAGATGATGCTCTG 1039  
 QY 119 TAGTGACAAAACAC 134  
 DB 1038 TAGGGGTCTAGACAC 1023

RESULT 3  
 ID Q50631/C  
 AC Q50631;  
 DT 02-JUN-1994 (first entry)  
 DE Brain factor-1.  
 KW Brain factor; BF-1; BF-2; BF-3; DNA binding domain;  
 KW brain transcription factor; diagnosis; tumour; cancer; probe;  
 KW telencephalon; ss.  
 OS Rattus rattus.  
 FH Key Location/Qualifiers  
 FT CDS 443..1885  
 FT misc-feature /\*tag- a  
 FT 926..1255  
 FT /\*tag- b  
 FT /note- "encoding DNA binding domain homology"  
 PN W09333430-A.  
 PD 25-NOV-1993.  
 PF 30-APR-1993; U04102.  
 PR 13-MAY-1992; US-882292.  
 PA (SLOK) SLOAN KETTERING INST CANCER.  
 PI Lai E, Tao W;  
 DR WPI: 93-386481/48.  
 DR P-PSDB: R44551.  
 PT Isolated, animal nucleic acids encoding brain transcription  
 PT factors - useful for diagnosis and treatment of abnormal brain  
 PT factor synthesis in tumour tissue from animals and utilised as  
 PT probe  
 PS Claim 1: Page 57-60; 96pp; English.  
 CC The brain factors are transcription factors. Abnormal expression of  
 CC BF-1 in telencephalon-derived tissue or tumour tissue can be  
 CC diagnosed. BF DNA or protein can be used to correct defective  
 CC synthesis of BF.  
 SO Sequence 2832 BP; 628 A; 868 C; 711 G; 625 T;

Query Match 12.3%; Score 28.2; DB 1; Length 2832;  
 Best Local Similarity 55.7%; Pred. No. 3.7;  
 Matches 54; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 96 TGCAGACAGCAGCGCTTGTAGTGCACAAAACACCAATGGCCCTACCTCCAAAT 155  
 DB 2604 TAGAAAAAATTAAGTACCGCTTGTAGTGCACACATACATGAAATATACACACAAAC 2545  
 QY 156 CCCAGCATTCCTAGCTTGCACCTTGAATACAGT 192  
 DB 2544 TGAAGCAATCCTAATTTGTCTTGGATTCAAT 2508

RESULT 4  
 ID T12460  
 AC T12460;  
 DT 21-APR-1996 (first entry)  
 DE Ovine adenovirus genome.  
 KW Ovine adenovirus; genome; sheep; inverted terminal repeat; E1A/B;  
 KW E1; E3; promoter; packaging signal; RNA processing signal;  
 KW plasmid pOAV100; cloning; vector; antigen; recombinant vaccine;  
 KW rotavirus; Trichostrogylus colubriformis; Teenia ovis;  
 KW Lucilia cuprina; antisense RNA; ribozyme; transgenic animal;  
 KW gene transfer; cyclic; ss.  
 OS Ovine adenovirus strain OAV287.  
 FH Key Location/Qualifiers  
 FT repeat\_region 1..46  
 FT /\*tag- a  
 FT /rpt-type= INVERTED  
 FT /note- "Inverted terminal repeat (claim 7)"  
 FT misc-feature 22139..22130  
 FT /\*tag- b  
 FT /note- "Non-essential region suitable for cloning"  
 FT misc-feature 26645..26646  
 FT /\*tag- c  
 FT /note- "Non-essential region suitable for cloning"  
 FT misc-feature 28644..28649  
 FT /\*tag- d  
 FT /note- "Non-essential region suitable for cloning"  
 PN W09603508-A1.  
 PD 08-FEB-1996.  
 PF 26-JUL-1995; AU0453.  
 PR 26-JUL-1994; AU-007101.  
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.  
 PI Both GW, Boyle DB, Vratil S;  
 DR WPI: 96-117053/12.  
 PT Viral vector derived from ovine adenovirus (OAV287) - used for  
 PT recombinant prodn. and delivery of non-adenoviral polypeptide(s) or  
 PT RNA molecules in host animals, esp. sheep.  
 PS Claim 1; Fig 1; 53pp; English.  
 CC The sequence represents the entire genome of ovine adenovirus OAV287,  
 CC a benign adenovirus isolated from sheep in Australia, and begins at  
 CC base 1 of the left-hand inverted terminal repeat. The sequence has  
 CC been determined by cloning BamHI restriction fragments representing  
 CC the entire genome and sequencing nested deletion mutants. Reading  
 CC frames have been tentatively identified, including E1A/B, E4 and E3.  
 CC The sequence contains promoters, inverted terminal repeats, viral  
 CC packaging signals and RNA processing signals. The entire genomic  
 CC clone is present in plasmid pOAV100. It is possible to introduce  
 CC recombinant DNA at non-essential sites, e.g. in the E3 region,  
 CC including a rotavirus VP/SC antigen, Trichostrogylus colubriformis  
 CC I7-kDa antigen, Teenia ovis 45w antigen or Lucilia cuprina PM95  
 CC antigen gene, or DNA encoding antisense RNA or a ribozyme. The  
 CC sequence may be used in construction of vectors to express  
 CC heterologous DNA in a grazing animal, e.g. a sheep. Many host  
 CC species have no pre-existing immunity to the virus, allowing  
 CC efficient infection, gene delivery and expression, and the  
 CC possibility of virus spread to a non-target host is minimized. The  
 CC virus may be used as a vector for recombinant vaccine production or  
 CC transgenic animal construction.

Sequence 29544 BP; 10167 A; 5365 C; 4562 G; 9450 T;

Query Match 12.2%; Score 28; DB 1; Length 29544;

Best Local Similarity 58.3%; Pred. No. 11;

Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 115 TTGTGAGTGACAAAACACCAAAATGGCTTCTTATCATCCAGGCAATGCTTAAGTT 174

DB 27819 TTTTCTTTTATTAACACAGACCATTCAGGCCAATTCAGTTATGTTTAATT 27878

QY 175 GGCACTTGAATAACAGCTGCA 198

DB 27879 TGCTACAGAACTGCAGACCA 27902

RESULT 5

V59086

ID V59086 standard; DNA; 3083 BP.

AC V59086;

DT 15-JAN-1999 (first entry)

DE Human ribonucleotide reductase R1 coding sequence.

KW Ribonucleotide reductase R1; human; tumorigenicity modulation;

KM neoplastic cell; tumour growth; tumour regression; therapy; ss.

OS Homo sapiens.

PN WO9841231-A1.

PD 24-SEP-1998.

PF 18-MAR-1998; CA0242.

PR 02-OCT-1997; US-060669.

PR 19-MAR-1997; US-040837.

PA (GENE-) GENESENSE TECHNOLOGIES INC.

PI Wright JA, Young AH;

DR WPI: 98-520956/44.

PT Use of ribonucleotide reductase R - for modulating the

PT tumorigenicity and metastatic potential of neoplastic cells in a

PS mammal, particularly for inhibiting tumour growth

CC Claim 6: Page 40-42; 56pp; English.

CC This sequence encodes the human ribonucleotide reductase R1 (RRR1). The

CC gene and the protein it encodes can be used in the methods of the

CC invention for modulating tumorigenicity of neoplastic cells in a mammal.

CC The RRR1 can slow tumour growth and cause tumour regression and reduction

CC of tumorigenicity and metastatic potential.

CC Sequence 3083 BP; 944 A; 599 C; 703 G; 837 T;

QY 13 GTGGACAGCAGCGTGGAAAGACACTTGTCTTATCATCCAGGCAAGCCTCC 72

DB 329 GGGGTACACACAGATGACATGCTGGCTGGAACACAGCTCAACTGACTACT 388

QY 73 CAATCTCTCTGGGATCCCACTGACAGAACAGAGTGGCTTTGAGTGACAAAC 132

DB 389 AAGCACCCTGATATCTATCTCTGCAAGCAGATGCTGTCTAATTGCAAGAA 448

QY 133 ACCAAA 139

DB 449 ACAAGA 455

RESULT 6

X40610

ID X40610 standard; CDNA; 468 BP.

AC X40610;

DT 18-JUN-1999 (first entry)

DE Human secreted protein 5 EST SEQ ID No: 210.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;

KM Human; gene therapy; chromosome mapping; signal peptide; prostate;

KW upstream regulatory sequence; cytokine activity; cell proliferation;

CC differentiation; haematopoiesis regulation; tissue growth regulation;

KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition; ds.

OS Homo sapiens.

PN WO9906550-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; IB1232.

PR 01-AUG-1997; US-905144.

PA (GENSE) GENSE.

PI Ductert A, Dumas Milne Edwards J, Lacroix B;

DR WPI: 99-153780/13.

P-PSDB: Y11888.

PT New isolated prostate-derived nucleic acids - used to develop

PT products which may have cytokine, immune regulatory, haematopoiesis

PT regulating, anti-inflammatory or tumour inhibition activity

CC Claim 1: Page 327-328; 67pp; English.

CC X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human

CC secreted proteins expressed in prostate, and encode the proteins given in

CC Y11716 to Y11993 respectively. The proteins given represent the signal

CC peptide and an N-terminal fragment of a secreted protein. The nucleic

CC acid sequences can be used for producing secreted human gene products.

CC They can also be used to develop products for diagnosis and therapy. The

CC proteins obtained may have cytokine activity, cell proliferation and

CC differentiation activity, haematopoiesis regulating activity, tissue

CC growth regulating activity, reproductive hormone regulating activity,

CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,

CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition

CC activity or other activities. The products can be used in forensic, gene

CC therapy and chromosome mapping procedures. The sequences can also be used

CC for obtaining corresponding promoter sequences. The nucleic acids

CC encoding the signal peptides can be used for directing extracellular

CC secretion of a polypeptide or the insertion of a polypeptide into a

CC membrane, or importing a polypeptide into a cell.

CC Sequence 468 BP; 89 A; 142 C; 116 G; 119 T;

QY 10 ATGTGACAGCAGCGTGGAAAGACACTTGTCTTATCATCCAGGCAAGCCT 69

DB 154 ATGTGACCGCTCACTGAGCCGCGGACACTCTTGTGTTTATGAGGACTATCAAGCTG 213

QY 70 CCCCAATCTCTCTGGGATCCCACTGACAGAACAGAGTG 112

DB 214 ACCCCAGGCTCAGCAGAGATGCTTACAGTGAATGAACAGTG 256

RESULT 7

Q90019/c

ID Q90019 standard; CDNA; 4213 BP.

AC Q90019;

DT 02-NOV-1995 (first entry)

DE Human betaglycan contg. transforming growth factor-beta binding site.

KW Betaglycan; transforming growth factor-beta; TGF-beta; binding site;

OS Homo sapiens.

PN key

FT Location/Qualifiers

FT cds 622..3168

FT /tag= a

PN WO9510610-A.

PD 20-APR-1995.

PF 14-OCT-1994; U11648.

PR 15-OCT-1993; US-138368.

PA (LJOL-) LA JOLLA CANCER RES FOUND.

PI Butzow R, Fukushima D, Ruoslahti E;

DR WPI: 95-178526/23.

P-PSDB: R74602.

PT New beta:glycan polypeptide(s) with TGF-beta binding activity -

PT useful for treating e.g. glomerulonephritis, diabetic

PT nephropathy, cirrhosis and fibrosis.

PS Disclosure; Fig 10: 63pp; English.

CC The nt and AA sequences of human betaglycan are given in Q90019

CC and R74602 and in Koren et al., Biochem. Biophys. Res. Comm.,





CC (encoded by the hmb gene) was encoded by ORF1. The protein can be used in vaccines for preventing or treating H. somnus infections, which cause thromboembolic meningo-encephalitis, septicaemia, CC arthritis and pneumonia in vertebrates.

CC See also 051081-6.

SO Sequence 2889 BP; 1094 A; 608 C; 549 G; 638 T;  
Query Match 11.8%; Score 27; DB 1; Length 2889;  
Best Local Similarity 60.0%; Pred. No. 9.6;  
Matches 45; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY 118 GTAGTGACAAAACACCAATGCGCTACCTCCCAATCCGACGCTGTAGCTTGGC 177  
DB 1336 GATTGTGACCTAAACCCACATTAGCCAACTTCCTCGCATGTGGCTGCTT 1395  
OY 178 ACTTGGAAATACAGT 192  
DB 1396 ATCTTGCCCTAATAGT 1410

RESULT 10  
X07291/c  
ID X07291 standard; cDNA: 1029 BP.  
AC X07291;  
DT 21-MAY-1999 (first entry)  
DE pig-tailed macaque HIV/SIV receptor protein Bonzo cDNA.  
KW SIV receptor; HIV receptor; Bonzo; translocation promoting agent;  
KM G protein coupled receptor; chemokine receptor; AIDS; infection;  
KW virus transmission; pig-tailed macaque; ds.  
OS Macaca nemestrina.  
PN W09030888 A1.  
PD 28-JAN-1999.  
PF 17-JUL-1998; U14857.  
PI 17-JUL-1997; US-896155.  
PA (UYNV) UNIV NEW YORK STATE.  
PI Deng H, Kewalramani VN, Littman DR, Unutmaz D;  
DR WPI: 99-132164/11.  
P-PSDB: W97785.  
PT New nucleic acid encoding the human translocation promoter Bonzo -  
PT used to screen for potential agents for treating acquired immune  
PT deficiency syndrome.  
PS Disclosure, page 81-82; 97pp; English.  
CC This cDNA clone codes for pig-tailed macaque Bonzo (see W97785), a  
CC novel HIV/SIV translocation promoting agent that acts in conjunction  
CC with CD4 to serve as a receptor for the entry into a cell of a virus  
CC having a specific viral envelope glycoprotein. Bonzo is a member of  
CC the 7-transmembrane G-protein coupled receptor family, is closely  
CC related to the chemokine receptor family, and is expressed in  
CC lymphoid tissues. The invention provides the amino acid sequences  
CC (see W97783-88) and DNA sequences (see X07289-94) of human,  
CC African green monkey and pig-tailed macaque Bonzo and Bob (brother  
CC of Bonzo) translocation promoting agents. These novel receptors  
CC were identified using an expression cloning strategy. They were  
CC found to be used by particular strains of HIV-2 and M-tropic HIV-1.  
CC The invention also includes: mammalian cells transfected with Bonzo  
CC and/or Bob and human CD4, which can be used to screen potential  
CC therapeutic agents and identify ligands; antibodies to Bonzo, which  
CC can be used therapeutically, e.g. as antagonists or to target  
CC toxins or radioisotopes to HIV-permissive cells; transgenic  
CC animals; and antisense and ribozyme molecules, which may also be  
CC used therapeutically, particularly expressed from a gene therapy  
CC vector.  
SQ Sequence 1029 BP; 230 A; 278 C; 236 G; 285 T;

OY 123 GCACAAAACACCAAAATGGCTTCCCAATCCGACGCTTGTAGCTTGACACTTG 182  
DB 553 TGCATGACGAATACACCATATGAGCTTGTCCAGTTAAAGACATTGCCATTAATTGGG 494

OY 183 GAAATACAGCTGTGACAGCA 201  
DB 493 GCAGGAAACACGACGAGGA 475

RESULT 11  
V77386/c  
ID V77386 standard; DNA: 376 BP.  
AC V77386;  
DT 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #3075.  
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
OS Staphylococcus aureus.  
PN EP-786519-A2.  
PD 30-JUL-1997.  
PF 07-JAN-1997; 100117.  
PI 05-JAN-1996; US-009861.  
PA (HUMAN) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
PI Rosen CA;  
DR WPI: 97-374922/35.  
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines.  
PS Claim 1; Page 2433; 3271pp; English.  
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S.aureus DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against S.aureus infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.  
SQ Sequence 376 BP; 102 A; 77 C; 60 G; 134 T;

Query Match 11.7%; Score 26.8; DB 1; Length 376;  
Best Local Similarity 58.2%; Pred. No. 4.9;  
Matches 46; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 69 TCCCAATCTCTCTCTGATCCAGTGACAGACAGCAAGTGTAGTGACAA 128  
DB 264 TGTCTATCAACGCTATGTGGCCCAANANATATATCTAGTGACTATGTGGAGAT 205  
OY 129 AAACACCAAAATGGCCTAC 147  
DB 204 ACAATATAAGATGTAAAC 186

RESULT 12  
Q13830  
ID Q13830 standard; DNA: 1920 BP.  
AC Q13830;  
DT 09-DEC-1991 (first entry)  
DE MSP gene in pUCRS (DSM 5803).

CC	N00006	give the sequences of cDNA corresp. to bps 1-98 and 1028-1129
CC	of N00004	respectively.
SO	Sequence	1742 BP; 589 A; 342 C; 408 G; 403 U;
Query Match		
Best Local Similarity 11.6%; Score 26.6; DB 1; Length 1742;		
Matches 56; Conservative 46.3%; Pred. No. 11;		
Matches 56; Conservative 6; Mismatches 59; Indels 0; Gaps 0;		
OY	78	CCTTCTCGGGATCCCACTGGACAGACAGGCTGGCTTTGTATGCGCAAAACACCAA 137
DB	1113	CCGUUCCAGGCAUACAGAAUUGCACAAGCAGAAAGCAUCGACAGACUACAAAGACCCCA 1172
OY	138	AATGSCCTACCTCCAAATCCAGCATGCTAAGGTTGCGACTTGGAAATACAGTCTCA 197
DB	1173	AUCCGCAUUAUACAGAAUACCGGAAAUUAAUAGACUUAUUGAGAAACCAACACGCA 1232
OY	198	A 198
DB	1233	A 1233
RESULT 14		
ID	X02665	standard; DNA; 8888 BP.
AC	X02665;	
DI	07-MAY-1999	(first entry)
DE	EP-892047	Seq ID 41.
KW	Semaphorin L; human; immunosuppressant; anti-inflammatory; gene therapy;	
KN	organ transplantation; inflammation therapy; immunotherapy; agonist;	
OS	immunomodulatory; antagonist; ss.	
PN	Homo sapiens.	
PD	EP-892047-A2.	
PE	20-JAN-1999.	
PF	06-JUL-1998; 112470.	
PR	11-FEB-1998; DE-005371.	
PT	09-JUL-1997; DE-029211.	
PA	(HMRI ) HOECHST MARION ROUSSEL DEUT GMBH.	
PI	Ensser A; Fleckenstein B;	
WP1	99-083564/08.	
PT	New semaphorin L proteins - used as immunosuppressants and	
PT	antiinflammatory agents in organ transplants, inflammation therapy,	
PT	immunotherapy and gene therapy	
PS	Disclosure; Page 105-111; 135pp; German.	
CC	This invention describes a novel human semaphorin L protein. This protein	
CC	or its encoding DNA are useful as immunosuppressants and/or	
CC	anti-inflammatory agents in organ transplantation, inflammation therapy,	
CC	immunotherapy and gene therapy. The DNA can be used to produce knock-out	
CC	or knock-in animals for research purposes. The proteins or DNA can be	
CC	used to search for the corresponding receptors or to screen for	
CC	immunomodulatory agonists or antagonists.	
SO	Sequence 8888 BP; 1773 A; 2359 C; 2758 G; 1998 T;	
Query Match		
Best Local Similarity 11.6%; Score 26.6; DB 1; Length 8888;		
Matches 59; Conservative 52.2%; Pred. No. 21;		
Matches 59; Conservative 0; Mismatches 54; Indels 0; Gaps 0;		
OY	14	TGGACAGCACCGTGGGAAGACACTTTGTTCTTATCTACTGGACACGACCTCCCC 73
DB	412	TGGGCTCAGCTCTTGATATGAGACCGCTTGAGCTAGGATGACTACCTCCTTCA 471
OY	74	AAATCTTCTCTGGGATCCCATGTGACAGAAAGCAAGTGGCTTTATATGAC 126
DB	472	GAGCCCACTGCTCTCATCTGCATGATGAGAAATATGTTGCTTCTCTTGGGGC 524
RESULT 15		
ID	013981/c	
AC	013981	standard; DNA; 1963 BP.
DI	02-JAN-1992	(first entry)
DE	Xylitol dehydrogenase - xyl-2 gene.	

KM XDH: xylose reductase: XR: ethanol: resistance: ss.  
OS Pichia stipitis CBS 5773 (DSM 5855).  
FH Key Location/Qualifiers  
FT cds 319..1407  
FT /\*tag= a  
FT /label= XDH

DE4009676-A.  
PD 02-OCT-1991.  
PE 26-MAR-1990; 009676.  
PR 26-MAR-1990; DE-009676.  
PA (RHEI-) RHEIN BIOTECH GES B.  
PI Strasser AMW, Hollenberg CP, Von Cliracy-Wantrup M, Kotter P;  
PI Amore R, Piontek, Hagedorn J;  
PI WPI: 91-296506/41.  
DR P-PSDB: R14181.  
PT New DNA encoding xylose reductase and xylitol dehydrogenase - and  
PT transformed yeast for prodn. of ethanol and biomass from xylose  
PT or recovery of oxidised NADP  
PS Disclosure: fig 2B: 50pp; German.  
CC XR and XDH were isolated and used to raise antipodies in mice and  
CC these used to screen a cDNA bank, in lambda gtl, derived from polyA-  
CC RNA of P.stipitis. XYL-1 clones (contg. two EcoRI fragments of 0.5  
CC and 0.6 kb) were identified, and their inserts placed in plasmid  
CC pT73-18U. The recombinant prods. were used to screen a genomic DNA  
CC bank prepd. in shuttle vector YEP13. Two plasmids were identified:  
CC pRI contg. the xyl-1 gene (Q13980) encoding XR and pDI contg. the  
CC xyl-2 gene encoding XDH.  
CC Microorganisms transformed with the xyl-1 and/or xyl-2 gene can ferment  
CC highly conc. carbohydrate solns. and are tolerant to EtOH, pH and  
CC temp.  
SQ Sequence 1963 BP; 513 A; 479 C; 410 G; 561 T;

Query Match 11.5%; Score 26.4; DB 1; Length 1963;  
Best Local Similarity 57.1%; Pred. No. 13;  
Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 23 CCGTGGGAAGGACACTTGTCTTATCACTGACAGCAGCCTCCCAATCCTC 82  
DB 219 CTGGAGAGATGCTCTTTATATATAAGCTGAGAACTATGTCCCGGTTATGT 160  
OY 83 TCTGGATCCAGTGGAGAGAGC 106  
DB 159 GCAGCAGCCGACAGAGCAGAGAGC 136

Search completed: August 6, 1999, 04:03:20  
Job time: 1552 sec

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:38:14 ; Search time 3268.17 Seconds

(without alignments)  
222.844 Million cell updates/sec

Title: US-09-049-696-10

Perfect score: 1 GGCACAGTGCATGTCGACAG.....AACCTGACCTGACTGTCA 229

Sequence: 1 GGCACAGTGCATGTCGACAG.....AACCTGACCTGACTGTCA 229

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_htg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_htg1:\*  
35: gb\_htg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
1	229	100.0	3311	11	AF039400		AF039400 Homo sapi

2	229	100.0	2826	42	AF127036	AF127036 Homo sapi
3	173	75.5	35278	11	AF039401	AF039401 Homo sapi
4	26	11.4	2937	12	AB017156	AB017156 Mus muscu
5	21	9.2	191356	42	AC005041	AC005041 Homo sapi
6	18	7.9	123322	7	ATF6118	AL022198 Arabidops
7	18	7.9	149298	9	AP000031	AP000031 Homo sapi
8	18	7.9	100000	9	AP000134	AP000134 Homo sapi
9	18	7.9	100000	9	AF000212	AF000212 Homo sapi
10	18	7.9	67008	9	AP000252	AP000252 Homo sapi
11	18	7.9	41506	11	AC004205	AC004205 Homo sapi
12	18	7.9	197915	11	AC005725	AC005725 Homo sapi
13	18	7.9	176153	11	AC006287	AC006287 Homo sapi
14	18	7.9	236822	34	D84394	D84394 Homo sapien
15	18	7.9	177294	35	AF048727	AF048727 Homo sapi
16	18	7.9	35491	42	AF048727	AF048727 Homo sapi
17	17	7.4	38916	1	MLCB2548	271163 Mycobacte
18	17	7.4	42861	1	MTCY339	L36547 Pseudomon
19	17	7.4	24489	1	PSEMT	X84072 S.pristinae
20	17	7.4	1806	1	SPDNAPTR	AF058919 Arabidops
21	17	7.4	91149	8	AF058919	AC007212 Arabidops
22	17	7.4	57550	8	ATAC007212	AB012723 Homo sapi
23	17	7.4	40850	9	HS370M22	Z82206 Human DNA
24	17	7.4	143747	9	HS370M22	AL031732 Human DNA
25	17	7.4	69017	10	HS41018	AL048738 Human DNA
26	17	7.4	122383	10	HS3944N18	AC003999 Human PAC
27	17	7.4	133457	11	AC003999	AC004636 Homo sapi
28	17	7.4	78293	11	AC004636	AC005553 Homo sapi
29	17	7.4	179651	11	AC005553	AC006257 Homo sapi
30	17	7.4	175818	11	AC006257	AF064865 Homo sapi
31	17	7.4	159424	11	AF064865	AC004682 Homo sapi
32	17	7.4	188134	11	HDAC004682	AC004155 Mus muscu
33	17	7.4	128026	12	AC004155	AC005960 Mus muscu
34	17	7.4	158414	12	AC005960	AL17502 Mouse mRNA
35	17	7.4	2453	12	MMECA39	G12349 Nile tilapi
36	17	7.4	335	14	G12349	G25428 human SRS E
37	17	7.4	417	14	G25428	G27691 human SRS S
38	17	7.4	42398	34	AC003095	AC003095 Homo sapi
39	17	7.4	27659	34	AC006629	AC006629 Caenorhab
40	17	7.4	299919	34	AC006712	AC006712 Caenorhab
41	17	7.4	215917	34	AC006764	AL031704 Homo sapi
42	17	7.4	40734	34	HS303A1	AL031654 Homo sapi
43	17	7.4	155116	34	HS303A1	AF034627 Phytomona
44	17	7.4	210	37	AF034627	
45	17	7.4	210	37	AF034627	

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	mRNA	PRI	15-DEC-1998
LOCUS		AF039400				
DEFINITION		Homo sapiens calcium-dependent chloride channel-1 (hclcal) mRNA, complete cds.				
ACCESSION		AF039400				
NID		94009457				
VERSION		AF039400.1	GI:4009457			
KEYWORDS		human.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS		Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,R.D., Fuller,C.M. and Pauli,B.U.				
TITLE		Genomic cloning, molecular characterization, and functional analysis of human hclcal, the first human member of the family of Ca2+-activated Cl- channel proteins				
JOURNAL		Genomics 54 (2), 200-214 (1998)				
MEDLINE		99047526				
REFERENCE		2 (bases 1 to 3311)				
AUTHORS		Gruber,A.D., Elble,R. and Pauli,B.U.				
TITLE		Direct Submission				

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hclcal1"  
352. 3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:g4009458"  
/db\_xref="GI:4009458"  
/translation="MGPFKSSVFILHLLEGALSLSLIQNNNGYEGIVADPNV  
EDTLIOQIDMYQASLYLEFANGKFFKNAVILIPETWKADYVRKLETKNA  
DVLVAESTPGNDEPTDOMNGCEKGERIHLIPDIAGKKALEYGPOGARVHEHA  
LWGVGEYNNDEKFTLSNGRIQAVCSAGITGVNKKCGGSCYTKCTFKVGL  
YKGCERFVLOSROTEKASIMFAOHVDSIVECEONHNKRNPNOKNLSSTWVI  
RDSDEFKTKTPMTQPNPFTSLIQIRYICVLDKSGSMATGNLNNOGOLF  
LQVLEGSWGAVTPSAHVOSELIQINSQSDRTAKRIPAAAGCTISIGLSRA  
FYIRKRPIDGSEIVLLTDGEDNTISGCFENYKQSAIIHYVALGPSAOELESK  
MTGGLQTVASDOVQNNGLIDAFGLSSGNGAVORSIQLESKGLTLQNSQMMGTIV  
DSTVGKDTFLITWTPOPIILMDPSGQKGFVVDKNTKNAVILQIPGAKVTWY  
SLOASQDTLITVTSRASNATLPITVTSKNTDSKFSPLVYANIRGASPIIRA  
SVTALIESVNGKVTLELDNGAGADATKDDGYSRYFTYDNGRYSVRVALGAVN  
AARRVTPQOOSGALYIPGWIENDTQNNPPREINKDDVQHKVCFRTSSGSEFAS  
DVPNAPIDLPFPQITDLKAEIHGSLINLTWAPDDDHGTAKHTIIRISTILD  
LRDNFNSLQVNTALIPKEANSEVFLFRPENTIFENGDTLFAIQAVKVDLKSFI  
SNIRVSLFIPQTPPETPSPDETSAPCPNINHSTIPGIHLIKMKWIGELQLSIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 229; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 4e-122;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACAGTGCATCTGGACAGCACCCTGGGAAAGACACTTTGTTCTATACCTGGACA 60  
|||||  
DB 1861 GGCACAGTGCATCTGGACAGCACCCTGGGAAAGACACTTTGTTCTATACCTGGACA 1920  
QY 61 ACGCAGCCTCCCAAAATCCTCTCGGATCCCAAGTGGACAGCAAGTGGCTTTGTA 120  
|||||  
DB 1921 ACGCAGCCTCCCAAAATCCTCTCGGATCCCAAGTGGACAGCAAGTGGCTTTGTA 1980  
QY 121 GTGACAAAACACCAAAATGGCTTACTCCAAATCCAGGCAATGCTAAGTGGCACT 180  
|||||  
DB 1981 GTGACAAAACACCAAAATGGCTTACTCCAAATCCAGGCAATGCTAAGTGGCACT 2040  
QY 181 TGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCTGACTGTA 229  
|||||  
DB 2041 TGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCTGACTGTA 2089

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID g4585468  
KEYWORDS AF127036.1 GI:4585468  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
2 (bases 1 to 2826)  
Aguel,M., and Culouscou,J.-M.  
Direct Submission  
Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrieres, Ruell-Malmaison 92500, France  
location/Qualifiers  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="CACCL1"  
5..2749  
/gene="CACCL1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PID:g4585469"  
/db\_xref="GI:4585469"  
/translation="MGPFKSSVFILHLLEGALSLSLIQNNNGYEGIVADPNV  
EDTLIOQIDMYQASLYLEFANGKFFKNAVILIPETWKADYVRKLETKNA  
DVLVAESTPGNDEPTDOMNGCEKGERIHLIPDIAGKKALEYGPOGARVHEHA  
LWGVGEYNNDEKFTLSNGRIQAVCSAGITGVNKKCGGSCYTKCTFKVGL  
YKGCERFVLOSROTEKASIMFAOHVDSIVECEONHNKRNPNOKNLSSTWVI  
RDSDEFKTKTPMTQPNPFTSLIQIRYICVLDKSGSMATGNLNNOGOLF  
LQVLEGSWGAVTPSAHVOSELIQINSQSDRTAKRIPAAAGCTISIGLSRA  
FYIRKRPIDGSEIVLLTDGEDNTISGCFENYKQSAIIHYVALGPSAOELESK  
MTGGLQTVASDOVQNNGLIDAFGLSSGNGAVORSIQLESKGLTLQNSQMMGTIV  
DSTVGKDTFLITWTPOPIILMDPSGQKGFVVDKNTKNAVILQIPGAKVTWY  
SLOASQDTLITVTSRASNATLPITVTSKNTDSKFSPLVYANIRGASPIIRA  
SVTALIESVNGKVTLELDNGAGADATKDDGYSRYFTYDNGRYSVRVALGAVN  
AARRVTPQOOSGALYIPGWIENDTQNNPPREINKDDVQHKVCFRTSSGSEFAS  
DVPNAPIDLPFPQITDLKAEIHGSLINLTWAPDDDHGTAKHTIIRISTILD  
LRDNFNSLQVNTALIPKEANSEVFLFRPENTIFENGDTLFAIQAVKVDLKSFI  
SNIRVSLFIPQTPPETPSPDETSAPCPNINHSTIPGIHLIKMKWIGELQLSIA  
"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 229; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 4e-122;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACAGTGCATCTGGACAGCACCCTGGGAAAGACACTTTGTTCTATACCTGGACA 60  
|||||  
DB 1514 GGCACAGTGCATCTGGACAGCACCCTGGGAAAGACACTTTGTTCTATACCTGGACA 1573  
QY 61 ACGCAGCCTCCCAAAATCCTCTCGGATCCCAAGTGGACAGCAAGTGGCTTTGTA 120  
|||||  
DB 1574 ACGCAGCCTCCCAAAATCCTCTCGGATCCCAAGTGGACAGCAAGTGGCTTTGTA 1633  
QY 121 GTGACAAAACACCAAAATGGCTTACTCCAAATCCAGGCAATGCTAAGTGGCACT 180  
|||||  
DB 1634 GTGACAAAACACCAAAATGGCTTACTCCAAATCCAGGCAATGCTAAGTGGCACT 1693  
QY 181 TGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCTGACTGTA 229  
|||||  
DB 1694 TGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCTGACTGTA 1742

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hclcal1) gene,  
DEFINITION complete cds.

ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Blble,R.C., Ji,H.U., Schreuer,K.D., Fuller,C.M. and Paul,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca<sup>2+</sup>-activated Cl<sup>-</sup> channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Blble,R. and Paul,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source 1. 35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1. .1617  
/gene="hCLCA1"  
1. .33522  
/gene="hCLCA1"  
767. .1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587. .1596  
/gene="hCLCA1"  
/note="putative"  
1618. .1708  
/gene="hCLCA1"  
/number=1  
join(1618. .1708,1962. .2383,6667. .6807,6979. .7126,9696. .9801,15454. .15631,18591. .18809,19783. .20010,22253. .22427,24522. .24628,26644. .26859,27447. .27708,28766. .28936,31840. .32079,32919. .33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962. .2383  
/gene="hCLCA1"  
/number=2  
join(2222. .2383,6667. .6807,6979. .7126,9696. .9801,15454. .15631,18591. .18809,19783. .20010,22253. .22427,24522. .24628,26644. .26859,27447. .27708,28766,31840. .32079,32919. .33510)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="pid:4009460"  
/db\_xref="gi:4009460"  
/translation="MGPRKSYFLLILHLLEGALSLSLIQLNNNGEIVAVIDNPV  
EDTLIQIKDWVTOASLYLFEATGKRFYFKVVALIIPETWTKADYVAPKLETKNA  
DLVAESTPPGNDDEPTQMGNGCEGKERIHLTPFIAKLAETGPOKAVHEMAH  
LWGVDEYNDEKFEYLSNGRIQAVKSGIGTGVNKKCGSCYTKRCPNKVTGL  
YKGCFFVLOSROTAKASIMFAOVDISYEPTQONHNKAPNKKONCNLSTWEVI  
RSEDEKFTPTPTLOPNTPTSLIOIGRTVGLVDKSGMATGRNRLNNOAGLFL  
LOTVFLGSGVGVTFDSAAHVSSELIQINSQSDPTLAKRLPAASGSGSTSGLSRA  
FTVIRKTFIDGSEIVLTDGEDNTISGCFNFKVSGATILITVALGPSAQLLELSK  
MTGGLQTYASDGVONNGLIDAFGALSSGNGAVSORSIQESGLTLQNSQAMNGVIV  
DSIVGKDTLFLITWTPOPIILMDPSGKOGGVVDKMTKNAVLIQIAGVGMKY  
SLQASQTLFLITVTSRASNATLPITVTSKTKNDISKPSPLVYVANIQSGAPILRA  
SVTALIESVNGKTVLELDNGADADDCGVSYFTTYDNGSVKVRALGGVN  
AARRVIPQGSALYIPGWIENDEIOWNPREINKDPOHQVQCESRSGSGSYVAS  
DVFNAPIPDLFPPGOITLDNAETIHGGSILNLWTAPGDYDHTGHTAKTITRISTILD

LRDRNESLQVNTALLIPKANSEVEFLKPENTPENGTDLFLIAQAVDKVLSKEI  
SNIAKVSLEFIPQPPETSPDETSPACPCPNHINSTIPGIIHLTKIMRWIGELQLSIA  
6667. .6807  
/gene="hCLCA1"  
/number=3  
6979. .7126  
/gene="hCLCA1"  
/number=4  
9696. .9801  
/gene="hCLCA1"  
/number=5  
15454. .15631  
/gene="hCLCA1"  
/number=6  
18591. .18809  
/gene="hCLCA1"  
/number=7  
19783. .20010  
/gene="hCLCA1"  
/number=8  
22253. .22427  
/gene="hCLCA1"  
/number=9  
24522. .24628  
/gene="hCLCA1"  
/number=10  
26644. .26859  
/gene="hCLCA1"  
/number=11  
27447. .27708  
/gene="hCLCA1"  
/number=12  
28766. .28936  
/gene="hCLCA1"  
/number=13  
31840. .32079  
/gene="hCLCA1"  
/number=14  
32919. .33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 75.5%; Score 173; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 9.4e-90;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGCACAGTATGATGTGACAGCACCGTGGAAAGACACCTTGTCTTATCACTGGACA 60  
|||||  
Db 26689 GGCACAGTATGATGTGACAGCACCGTGGAAAGACACCTTGTCTTATCACTGGACA 26748  
QY 61 AGCGACCTCCCAATCTCTCTGTGGATCCCAAGTGGACAGAAAGCAAGAGCTTGTA 120  
|||||  
Db 26749 AGCGACCTCCCAATCTCTCTGTGGATCCCAAGTGGACAGAAAGCAAGAGCTTGTA 26808  
QY 121 GTGGACAAAACACCAAAATGGCTCACTCAATCCAGCATTTGCTAAGT 173  
Db 26809 GTGGACAAAACACCAAAATGGCTCACTCAATCCAGCATTTGCTAAGT 26861  
RESULT 4  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS AB017156  
DEFINITION Mus musculus gob-5 mRNA, complete cds.  
ACCESSION AB017156  
NID 93721911  
VERSION AB017156.1 GI:3721911  
KEYWORDS GOB-5.  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus

REFERENCE 1 (sites)  
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice  
JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
MEDLINE 99160866  
REFERENCE 2 (bases 1 to 2937)  
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) to the DBJ/EMBL/GenBank databases, Toru Komiyama, ERATO, JST, Hirohashi Cell Configuration Project, 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan (E-mail: tkomihncp.jst.go.jp, Tel: 81-298-47-7563, Fax: 81-298-47-5226)  
FEATURES  
source Location/Qualifiers  
1..2937  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15..2756  
/gene="gob-5"  
15..2756  
/gene="gob-5"  
/codon\_start=1  
/protein\_id="BAA3743.1"  
/db\_xref="PID:d1034712"  
/db\_xref="PID:g3721912"  
/db\_xref="GI:3721912"  
/translation="MESLSPVFLILHLEGVSESLIQNNNGYEGIVAIIDHDVPEDEALIOHKDMVQASPYLEATGKRPFRNVAILPESKARPEYRPLETKFKNADLVSTSPLENDDEPYELGACGEGKRIHLTPDLFLAGKLTQGPQDRFVEMAFRMGVFENYNNDEKFEYLSKKPOAVGSAATLTKNQVRCOGSGCITNGKCVIRWTGLKNCVFPVPHONERKASIMFNQINSVFECFENKNNQADNRQCRCLBSTWELQSEDFKQTPMTRQPPAPTFSLQIQRYCVCLVDSGSKMLNDNRNRNMQSREFLLOVEQSGWGMVTFDSAAVQSELRKOLNSGADDLIKHLPVVSAGTISCGLRTAFYIKKRPYDGESEIVLITDGEDNTISSCDLVKQSAIITHVALGPAARKEQLSKMTGGLQYSSDQVONGLVDAFALSSGNALIAQHSIQLESRGYNLONNMNGSVIVDSYSGKDTLLFTWTTPPTLFIWDPGVONGIILDTTKVAILQVGTAKGFMKYSQASSQTLTLTVSRASATLPIPTVPPVNNKTKGPEPPTVYASIRGASPIILRASVRLISVNGKTYTLELLDNGAGADTKNDGYSRFTAFDANGRSYVYALGCVTSDRQAPKPNRATIDGWLDEGEVRNRPPEPSTYQDQLCSRTSSGSEVAINVPAAPLPDLPPOITDLKASIQGNLVNLTWAPGDYDHGRASNTIIMKSVIIDLHDHNTSLQVNTGLIPKEASSEIEFEELGNTGFIATIQVADMSNLKSEISNIRSVFIRPAOEPPIEDSTPPCPDISINSTIPGIVHLKIMKWLQEMQVTLGLH"

BASE COUNT 860 a 718 c 693 g 666 t

ORIGIN

Query Match 11.4%; Score 26; DB 12; Length 2937;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 40 TTGTTCTTATCAGCTGGACACGCA 65  
|||||  
Db 1566 TTGTTCTTATCAGCTGGACACGCA 1591

RESULT 5  
AC005041 191356 bp DNA PRI 24-MAR-1999  
LOCUS Homo sapiens clone NH0523H20, complete sequence.  
DEFINITION AC005041  
AC005041  
NTD 94508118  
VERSION AC005041.2 GI:4508118  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 191356)  
AUTHORS Waterston, R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 191356)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 3 (bases 1 to 191356)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT On Mar 24, 1999 this sequence version replaced gi:3309088.  
FEATURES  
source Location/Qualifiers  
1..191356  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NH0523H20"

BASE COUNT 46280 a 47882 c 48007 g 49187 t

ORIGIN

Query Match 9.2%; Score 21; DB 42; Length 191356;  
Best Local Similarity 100.0%; Pred. No. 0.087;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 CAGCTCCCAATCCTCTC 84  
|||||  
Db 4335 CAGCTCCCAATCCTCTC 4335

RESULT 6  
ATF6118 122322 bp DNA PIN 25-MAR-1999  
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone F6118 (ESSAII project).  
DEFINITION ALO22198  
ACCESSION 92980757  
NID ALO22198.1 GI:2980757  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Caprales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 122322)  
AUTHORS Bevan, M., Benes, V., Reemann, S., Borikova, D., Ansgore, W., Honelisel, J., Wewes, H.W., Mayer, K. and Schueller, C.  
TITLE Unpublished  
JOURNAL EU Arabidopsis sequencing project.  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (08-APR-1998) MIPs, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk

FEATURES  
source Location/Qualifiers  
1..122322  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
199..921  
/gene="F6118.10"  
/complement(199..347)  
/gene="F6118.10"  
/number=1  
complement(join(199..347,404..548,631..921))



/gene="F6118.10"  
 /note="contains EST gb:122575, T22317"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CA18185.1"  
 /db\_xref="PID:e1283542"  
 /db\_xref="PID:g445153"  
 /db\_xref="GI:445153"  
 /translation="MAOQEGAVYKEGPNDSSAAVAATATATADSYKKONQFFSR  
 IANJGFRRGDDEFKRLQYISREKATVYSRKRRSISRKLRLNIYSSVFEIIAAG  
 YALLTPREDIDRMRKSRFRILPFIPLPVSALATSISYFSKMSPLILCKNCIYD  
 RRDKTLKRLRERLAKINELKERTNIYTTOOLQ"  
 complement(348, .403)  
 /gene="F6118.10"  
 /number=1  
 complement(404, .548)  
 /gene="F6118.10"  
 /number=2  
 complement(549, .630)  
 /gene="F6118.10"  
 /number=2  
 complement(631, .921)  
 /gene="F6118.10"  
 /number=3  
 1570, .3411  
 /gene="F6118.20"  
 /number=1  
 1570, .3411  
 /gene="F6118.20"  
 1570, .3411  
 /gene="F6118.20"  
 /note="similarity to various predicted proteins,  
 Arabidopsis thaliana"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CA18186.1"  
 /db\_xref="PID:e1283543"  
 /db\_xref="PID:g2980759"  
 /db\_xref="GI:2980759"  
 /db\_xref="SPTREMBL:065543"  
 /translation="MSRLSSRLNLELNKILKGLVSDFYDEALRLKKIKHSLGNG  
 FTALIPSYIKACAFQOEPLFLGAOLHCLCLAGADCDIVSNSLSIMAFSRVAVR  
 KVFDEMLRVDVSYCSINSCCODGLYEAKLLEKVFYGFIPKSEIVASLALCTR  
 MSSSKVARMFHAVLVDERMQESVLSIALVDMYLKEDDRAAFHYEDQVEVENES  
 WTAMSGVANONEMGVDFLRAMORENLRENRVTLISVLPACVELNYSVLEIENG  
 FSRFGCHADERLTPAFMTMCRGCVNSLSVLEPESKVRPVYVWSSMISYATGPC  
 SEVNNLLNOMREGIEANSVTLIVASACTSITLSLEFASVHSOLKCGFSHLLLN  
 ALIDWYACGSLSSAREFVETLTERDILVSSMSINAYGLHGGSEALEIFGKMGKH  
 EVDVAAPLALISACNHAGLVEAQITFQOAKYHNPVYLEHYACYINLGFQGLIDDA  
 FEVILNPMKPSARISLSLSCETHGLDVAQKIIANELKSRPDNPANVLLSKLKH  
 TESGNYHAEEVRRRMQRKLKNGCFEKIEPELQIEDYQKSNSP1"  
 4648, .6666  
 /gene="F6118.30"  
 complement(john(4648, .5128, 6668, .6966))  
 /gene="F6118.30"  
 /note="similarity to TINV, Arabidopsis thaliana,  
 PATCHX:E218696"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CA18187.1"  
 /db\_xref="PID:e1283544"  
 /db\_xref="PID:g2980760"  
 /db\_xref="GI:2980760"  
 /db\_xref="SPTREMBL:065544"  
 /translation="WYITGNOKVAAGIGIRSVKWTYTHGLAINTVTDLTPFNWIVGCI  
 RDNRKGNKGLLEDEGHGMWDCLRIDIVHSSLKRESEAPQOLIEKQTVSNPILAV  
 RKRSGKNSVSELEPGKTRGIRMLGSDVPEKAAAYALALCISGEKGVNFPPDKRP  
 OLPESSVPLSLDIQITATYVASSVVPVSHATLPATTPVSEVPASSVDSASTBI  
 TEMDEYVLPDARATESIFSYVEDQLDSFLMMDIDWNNLI"  
 complement(4648, .5128)  
 /gene="F6118.30"  
 /number=1

Intron	complement(5129..6667) /gene="F6118.30" /number=1 complement(6668..6966) /gene="F6118.30" /number=2 7019..7762 /gene="F6118.40" complement(join(7019..7195,7282..7591,7632..7762)) /gene="F6118.40" /note="similarity to LipB gene, <i>Prochlorothrix hollandica</i> , PATCX:G1399931" /codon_start=1 /product="putative protein" /protein_id="CA18188.1" /db_xref="pid:e1283545" /db_xref="pid:g2980761" /db_xref="gi:2980761" /db_xref="SPTRMBL:O65545" /translation="MELNGVELVSGIHHHRTNARNRLVRSVKILNSGNEIPRK HFNELTPNNHQFCECLDFYDKLVPRKAMSWQSKIIVEEKTLLIDRNDCAADTVLL LOHSVYMGASTEDYLNFDIKDAFENVYRTERGGVYTPGOLWVYPIINLNEE MDLHYLTMLEIYIRVLSSTFSIKASHDLGLTGWGMHSING"
exon	complement(7019..7195) /gene="F6118.40"
Intron	/number=1 complement(7196..7281) /gene="F6118.40" /number=1 complement(7282..7591) /gene="F6118.40" /number=2 complement(7592..7631) /gene="F6118.40" /number=2 complement(7632..7762) /gene="F6118.40"
exon	/number=3 8247..9367 /gene="F6118.50" join(8247..8864,9004..9089,9190..9367) /gene="F6118.50"
CDS	/codon_start=1 /product="hypothetical protein" /protein_id="CA18189.1" /db_xref="pid:e1283546" /db_xref="pid:g2980762" /db_xref="pid:g2980762" /db_xref="gi:2980762" /db_xref="SPTRMBL:O65546" /translation="MVLCICLSFSSQNPDESSSSLLRYKPCDSISLAKGRKKLNR FVPSKSNKSTGNNKRRRMORPFEDDGNVLGRDDIYDETLERAKDDMSDE EKFLWRKRAIAYELIREGDEIGDSGVADYTKKMDWDIYSDSPFESWTDSSGS DNLEEDLRTIPDGGLYKMRDNLVGAEEEDILIEDTIFRAASKSKAFVAILILPW ALDPLAHYLLTPFLDRYKVTPLAQTLDVRRNQKLEWAKELNREKARYLREVEIGK SPPLSDDLWEMNGKA" 8247..8864 /gene="F6118.50" /number=1 8865..9003 /gene="F6118.50" /number=1 9004..9089 /gene="F6118.50" /number=2 9090..9189 /gene="F6118.50" /number=2 9190..9367 /gene="F6118.50" /number=3 10508..10890 /gene="F6118.60" join(10508..10621,10870..10890)
CDS	



```

RESULT 9
AP000212 100000 bp DNA PRI 14-MAY-1999
LOCUS Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
DEFINITION clone f43d11-119b8, segment 10/12, complete sequence.
ACCESSION AP000212
NID 94827160
VERSION AP000212.1 GI:4827160
KEYWORDS Homo sapiens DNA.
SOURCE HTG.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
TITLE Homo sapiens 1,109,292bp genomic DNA of 21q22.1 (REGION:
D21S226-AML CLONE RANGE: f43d11-119b8)
JOURNAL Published Only in Database (1999) In press
2 (bases 1 to 100000)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
TITLE Direct Submission
Submitted (10-MAY-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-42-778-9923, Fax:81-42-778-9924)
E. coli transposon insertion: The present data does not contain E.
colli transposon sequences which integrated in the
original/previous sequences. We determined the boundary between
the insertion and genomic sequences experimentally. removed the
insertion sequences, reconstituted the present data. The sequencing
project is supported by Japan Science Technology Corporation (JST)
and The Institute of Physical and Chemical Research (RIKEN).
FEATURES
Source
1.100000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
BASE COUNT 27831 a 22543 c 23121 g 26505 t
ORIGIN
Query Match 7.9%; Score 18; DB 9; Length 100000;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 178 ACTTGGAATACAGCTG 195
|||||
Db 96791 ACTTGGAATACAGCTG 96808

```

```

REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
TITLE Direct Submission
Submitted (13-MAY-1999) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-42-778-9923, Fax:81-42-778-9924)
The sequence is a part of the data (ACCESSION No. AP000203 -
AP000214).
COMMENT The sequencing project is supported by Japan Science Technology
Corporation (JST) and The Institute of Physical and Chemical
Research (RIKEN).
FEATURES
Source
1.67008
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="T650"
/map="21q22.1"
BASE COUNT 18394 a 14838 c 15419 g 18357 t
ORIGIN
Query Match 7.9%; Score 18; DB 9; Length 67008;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 178 ACTTGGAATACAGCTG 195
|||||
Db 59640 ACTTGGAATACAGCTG 59657

```

```

RESULT 11
AC004205
LOCUS Homo sapiens clone UMGc:y3c071 from 6p21, complete sequence.
DEFINITION AC004205
ACCESSION 93927854
NID AC004205.1 GI:3927854
VERSION HTG.
KEYWORDS Homo sapiens
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Gullandaux,T., Janer,M., Wong,G.K., Spies,T. and Geraghty,D.E.
TITLE The complete genomic sequence of 424,015 bp at the centromeric end
of the HLA class I region: gene content and polymorphism
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (16), 9494-9499 (1998)
98356185
REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
Contact: Daniel E. Geraghty (geraghty@fhcr.org)
2 (bases 1 to 41506)
Geraghty,D.E. and Olson,M.V.
TITLE Direct Submission
Submitted (23-FEB-1998) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 41506)
Geraghty,D.E. and Olson,M.V.
REFERENCE
AUTHORS Submitted (25-NOV-1998) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
On Nov 25, 1998 this sequence version replaced gi:2905855.
Overlapping Sequences:
5': UMGc:y3c028 (Genbank Accession: AC004210)
3': UMGc:y3c018 (Genbank Accession: AC004204)
-----

```

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 76.6%  
DS or two chemistry coverage: 99.9%  
Single stranded regions: 2

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

Map	Seq	HandIII	Seq	Map	Seq	NotI	Seq
6493.25	6534.00	2906.67	2931.00	965.88	958.00		
10020.00	9916.00	2230.67	2225.00	4984.38	4948.00		
756.25	755.00	3744.62	3714.00	675.00	674.00		
1474.00	1475.00	1451.75	1453.00	3790.75	3762.00		
2503.75	2519.00	1023.25	1015.00	3344.25	3327.00		
5066.00	5111.00	2795.86	2795.00	1157.75	1152.00		
2344.00	2361.00	2232.50	2258.00	2747.00	2747.00		
1916.00	1938.00	5171.00	5176.00	677.67	682.00		
5301.00	5219.00	1910.00	1924.00	677.67	667.00		
981.00	972.00	1441.67	1447.00	4056.00	4038.00		
1253.00	1266.00	1093.00	1095.00				
		9094.50	8842.00				
		841.00	831.00				

## FEATURES

source 1. 41506  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21"  
/clone="CGM1:B38D3"  
/sub\_clone="DMC:y3c071"  
/cell\_line="CGM1"  
/clone\_lib="Wash U YAC library"  
10. 227  
repeat\_region  
/rpt\_family="Alu"  
1270. 1504  
repeat\_region  
/rpt\_family="MER25"  
complement(1837. 2123)  
/rpt\_family="Alu"  
2135. 2300  
repeat\_region  
/rpt\_family="L1"  
complement(2581. 3065)  
/rpt\_family="MER9"  
3281. 4199  
repeat\_region

repeat\_region /rpt\_family="L1"  
complement(4733. 4871)  
repeat\_region /rpt\_family="LRR10"  
11921. 12186  
repeat\_region /rpt\_family="Alu"  
complement(13151. 13371)  
repeat\_region /rpt\_family="LRR10"  
13442. 13560  
repeat\_region /rpt\_family="L1"  
complement(13919. 19370)  
repeat\_region /rpt\_family="L1"  
15112. 15394  
repeat\_region /rpt\_family="Alu"  
15524. 15820  
repeat\_region /rpt\_family="Alu"  
17260. 17551  
repeat\_region /rpt\_family="Alu"  
complement(19213. 19447)  
repeat\_region /rpt\_family="MER25"  
complement(19538. 19648)  
repeat\_region /rpt\_family="L1"  
20344. 20620  
repeat\_region /rpt\_family="Alu"  
21823. 22104  
repeat\_region /rpt\_family="Alu"  
23830. 23893  
repeat\_region /rpt\_family="MLT1"  
25591. 25687  
repeat\_region /rpt\_family="Alu"  
complement(26880. 27051)  
repeat\_region /rpt\_family="MER3"  
30616. 30864  
repeat\_region /rpt\_family="Alu"  
complement(34535. 34834)  
repeat\_region /rpt\_family="Alu"  
complement(35723. 35832)  
repeat\_region /rpt\_family="MIR"  
38630. 38916  
repeat\_region /rpt\_family="Alu"  
38943. 39423  
repeat\_region /rpt\_family="L1"  
BASE COUNT 12317 a 8781 c 8500 g 11908 t  
ORIGIN

Query Match 7.9% Score 18; DB 11; Length 41506;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 TCCTCTCTGGGATCCCA 94  
DB 40014 TCCTCTCTGGGATCCCA 40031

RESULT 12  
AC005725 AC005725 197915 bp DNA PRI 02-DEC-1998  
LOCUS Homo sapiens chromosome 17, clone hRPX.680\_C.21, complete sequence.  
DEFINITION AC005725  
ACCESSION 93947428  
NID AC005725.1 GI:3947428  
VERSION HTG.  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 197915)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 17, clone hRPX.680\_C.21  
JOURNAL Unpublished  
2 (bases 1 to 197915)  
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Batlowin,J., Barne,N., Beckery,R., Benn,J., Boutwell,C.,

Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., Corliss, D., Depayre, E., Deron, K.,  
Donelan, L., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R.,  
Gage, D., Gardyna, S., Geralt, K., Grant, G., Hagos, B., Heatford, A.,  
Hereana, L., Horton, L., Howland, J. C., Jacot, L., Jones, C., Kann, L.,  
Karatas, A., Lehoczy, J., Macdonald, P., Margolis, N., McEwan, P.,  
McGuirk, A., McKernan, K., Meidrim, J., Molla, M., Morris, M., Morrow, J.,  
Mychalack, J., Nahf, R., Naylor, J., Nilotf, M., O'Connor, T.,  
O'Donnell, P., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Roy, A.,  
Severy, P., Stange-Thomann, N., Stillewell, J., Stojanovic, N., Stone, C.,  
Subramanian, A., Testfaye, S., Tichovolsky, N., Torrueila-Miller, I.,  
Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Wyman, D.,  
Ye, W. J., Zhao, J. and Zody, M.

Direct Submission  
Submitted (29-SEP-1998) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 197915)

REFERENCE  
AUTHORS  
Birken, B., Baldwin, J., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Boutwell, C.,  
Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., Corliss, D., Depayre, E., Devon, K., Dewar, K.,  
Donelan, L., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R.,  
Gage, D., Gardyna, S., Geralt, K., Grant, G., Hagos, B., Heatford, A.,  
Hereana, L., Horton, L., Howland, J. C., Jacot, L., Jones, C., Kann, L.,  
Karatas, A., Lehoczy, J., Macdonald, P., Margolis, N., McEwan, P.,  
McGuirk, A., McKernan, K., Meidrim, J., Molla, M., Morris, M., Morrow, J.,  
Mychalack, J., Nahf, R., Naylor, J., Nilotf, M., O'Connor, T.,  
O'Donnell, P., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Roy, A.,  
Severy, P., Stange-Thomann, N., Stillewell, J., Stojanovic, N., Stone, C.,  
Subramanian, A., Testfaye, S., Tichovolsky, N., Torrueila-Miller, I.,  
Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Wyman, D.,  
Ye, W. J., Zhao, J. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Only the last 197915 base pairs of this clone are being submitted.  
The remainder overlaps accession number AC005548 (WICGR project  
L361).

FEATURES  
SOURCE  
1. 197915  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="hrpk.680.C.21"  
/clone.lib="RPCT-11 human BAC library"  
/map="17"  
/chromosome="17"  
complement(251..355)  
/rpt\_family="MIR"  
687..969  
/rpt\_family="AluSc"  
2640..2661  
/rpt\_family="AT-rich"  
4789..5192  
/rpt\_family="MT1B"  
complement(5330..5795)  
/rpt\_family="MT1H"  
complement(7650..7676)  
/rpt\_family="(TGA)n"  
complement(7722..8030)  
/rpt\_family="LFR16A"  
complement(8051..8475)  
/rpt\_family="L2"  
complement(8580..8676)  
/rpt\_family="MIR"  
complement(8866..8935)  
/rpt\_family="MT1I"  
8936..8995  
/rpt\_family="(TAGA)n"

repeat\_region complement(8998..9054)  
/rpt\_family="(TAGA)n"  
repeat\_region complement(9057..9361)  
/rpt\_family="MT1I"  
repeat\_region 9391..9691  
/rpt\_family="AluSc"  
9756..9824  
/rpt\_family="polyurine"  
complement(9825..9847)  
/rpt\_family="MT1J"  
9848..9966  
/rpt\_family="(GAA)n"  
complement(9967..10060)  
/rpt\_family="MT1J"  
10947..11128  
/rpt\_family="L2"  
11187..11570  
/rpt\_family="L2"  
complement(12968..13015)  
/rpt\_family="(CGG)n"  
13110..13159  
/rpt\_family="GC-rich"  
complement(13396..13420)  
/rpt\_family="(CGG)n"  
complement(13586..13654)  
/rpt\_family="(CGG)n"  
13667..13758  
/rpt\_family="GC-rich"  
13884..13927  
/rpt\_family="GC-rich"  
13973..14023  
/rpt\_family="GC-rich"  
18346..18648  
/rpt\_family="AluY"  
complement(20108..20409)  
/rpt\_family="AluY"  
complement(20687..20805)  
/rpt\_family="MIR"  
21194..21995  
/rpt\_family="AluY"  
complement(21620..21682)  
/rpt\_family="(CA)n"  
complement(23260..23416)  
/rpt\_family="MIR"  
26356..26614  
/rpt\_family="MIR"  
27378..27513  
/rpt\_family="L2"  
27644..27727  
/rpt\_family="purine-rich"  
complement(28209..28257)  
/rpt\_family="L2"  
complement(28225..28485)  
/rpt\_family="MIR"  
28549..28847  
/rpt\_family="AluSc"  
29715..30026  
/rpt\_family="AluSg"  
30027..30068  
/rpt\_family="polyurine"  
30328..30353  
/rpt\_family="AT-rich"  
31266..31388  
/rpt\_family="(TGA)n"  
31474..31520  
/rpt\_family="purine-rich"  
31537..31784  
/rpt\_family="MIR"  
complement(33067..33186)  
/rpt\_family="purine-rich"  
complement(33831..34379)  
/rpt\_family="L1MB6"  
complement(34377..34527)  
repeat\_region



```

repeat_region complement(10287, .10373)
repeat_region /rpt_family="MIR"
repeat_region complement(10466, .10509)
repeat_region /rpt_family="MER5A"
repeat_region complement(10541, .11062)
repeat_region /rpt_family="MER74B"
repeat_region /rpt_family="MER58B"
repeat_region /rpt_family="MER58B"
repeat_region complement(13205, .13513)
repeat_region /rpt_family="ALUSX"
repeat_region complement(15208, .15655)
repeat_region /rpt_family="LTR7"
repeat_region /rpt_family="MER91A"
repeat_region /rpt_family="MER91A"
repeat_region /rpt_family="MER91A"
repeat_region complement(17844, .18227)
repeat_region /rpt_family="L2"
repeat_region complement(18257, .18820)
repeat_region /rpt_family="MER39"
repeat_region /rpt_family="MER39"
repeat_region /rpt_family="MLT1C"
repeat_region /rpt_family="L2"
repeat_region complement(20224, .20712)
repeat_region /rpt_family="L1PB1"
repeat_region complement(20713, .21010)
repeat_region /rpt_family="ALUY"
repeat_region complement(21011, .22272)
repeat_region /rpt_family="L1PB1"
repeat_region /rpt_family="MER69A"
repeat_region /rpt_family="MER69A"
repeat_region /rpt_family="L2"
repeat_region /rpt_family="L2"
repeat_region /rpt_family="MLT1"
repeat_region /rpt_family="L2"
repeat_region complement(23537, .23755)
repeat_region /rpt_family="MIR"
repeat_region /rpt_family="MIR"
repeat_region /rpt_family="AT_rich"
repeat_region complement(24184, .24491)
repeat_region /rpt_family="ALUSX"
repeat_region complement(25266, .25337)
repeat_region /rpt_family="LTR16A1"
repeat_region complement(25338, .25722)
repeat_region /rpt_family="MSTA"
repeat_region complement(25723, .26049)
repeat_region /rpt_family="LTR16A1"
repeat_region /rpt_family="ALUJo"
repeat_region /rpt_family="ALUJo"
repeat_region /rpt_family="GA-rich"
repeat_region /rpt_family="MER96"
repeat_region complement(27280, .27341)
repeat_region /rpt_family="MLT1H"
repeat_region /rpt_family="L2"
repeat_region /rpt_family="L2"
repeat_region /rpt_family="TCCA"n"
repeat_region /rpt_family="TCCA"n"
repeat_region complement(29515, .29794)
repeat_region /rpt_family="ALUSX"
repeat_region /rpt_family="TTC"n"
repeat_region complement(30898, .31206)
repeat_region /rpt_family="ALUJb"
repeat_region complement(33150, .33341)
repeat_region /rpt_family="MIR"
repeat_region /rpt_family="MIR"

```

```

repeat_region /rpt_family="TTC"n"
repeat_region complement(33698, .33807)
repeat_region /rpt_family="MER102"
repeat_region /rpt_family="AT_rich"
repeat_region complement(36166, .36292)
repeat_region /rpt_family="MIR"
repeat_region /rpt_family="AT_rich"
repeat_region complement(38233, .38355)
repeat_region /rpt_family="FLAM_A"
repeat_region /rpt_family="L2"
repeat_region complement(40466, .40588)
repeat_region /rpt_family="MIR"
repeat_region /rpt_family="MIR"
repeat_region complement(42147, .42440)
repeat_region /rpt_family="ALUSg"
repeat_region complement(42454, .42761)
repeat_region /rpt_family="ALUY"
repeat_region /rpt_family="MADE1"
repeat_region /rpt_family="AT_rich"
repeat_region /rpt_family="ALUJo"
repeat_region /rpt_family="ALUJo"
repeat_region complement(46671, .46671)
repeat_region /rpt_family="MER58A"
repeat_region complement(46698, .46827)

Query Match 7.9%; Score 18; DB 11; Length 176153;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 CAACCTGACCTGACT 225
Db 149508 CAACCTGACCTGACT 149525

RESULT 14
LOCUS DB4394 236822 bp DNA HTG 16-SEP-1997
DEFINITION Homo sapiens chromosome 6 clone Y109 map 6p21.3, WORKING DRAFT
ACCESSION DB4394
NID 92401262
VERSION DB4394.1 GI:2401262
KEYWORDS HTG; HTGS-PHASE2;
SOURCE Homo sapiens cell_line:CGM1 DNA, clone_lib:T.Imai and M.V.Olson
clone:Y109.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 236822)
INNOV H.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1996) to the DBJ/EMBL/Genbank databases.
Hidetoshi Inoko, Tokai University School of Medicine, Molecular
Life Science; Bohseida, Isehara Kanagawa 259-11, Japan
(E-mail:hinoke@is.lcc.u-tokai.ac.jp, Tel:0463-93-1121,
Fax:0463-94-8884)
2 (bases 1 to 236822)
REFERENCE
AUTHORS Mizuki,N., Ando,H., Kimura,M., Ohno,S., Miyata,S., Yamazaki,M.,
Tashiro,H., Watanabe,K., Ono,A., Taguchi,S., Sugawara,C.,
Fukuzumi,Y., Sato,M., Okumura,K., Goto,K., Ishihara,M., Kikuchi,Y.,
Shina,T., Ando,A., Ikemura,T. and Inoko,H.
Nucleotide sequence analysis of the HLA class I region spanning the
237 kb segment around the HLA-B and -C genes
Unpublished (1996)
JOURNAL Mizuki,N., Ando,H., Kimura,M., Ohno,S., Miyata,S., Yamazaki,M.,
REFERENCE Tashiro,H., Watanabe,K., Ono,A., Taguchi,S., Sugawara,C.,
AUTHORS

```

Fukuzumi,Y., Okumura,K., Goto,K., Ishihara,M., Nakamura,S.,  
Yonemoto,J., Kikuchi,Y.Y., Shilina,T., Chen,L., Ando,A., Ikemura,T.  
and Inoko,H.  
Nucleotide sequence analysis of the HLA class I region spanning the  
237-Kb segment around the HLA-B and -C genes  
97321045  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1..236822  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="CGM1"  
/chromosome="6"  
/clone="Y109"  
/clone\_11b="T. Imai and M.V.Olson"  
/map="6p21.3"  
BASE COUNT 66997 a 52352 c 52463 g 65010 t  
ORIGIN  
Query Match 7.9%; Score 18; DB 34; Length 236822;  
Best Local Similarity 100.0%; Pred.No. 4.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 77 TCCTCTCTGGGATCCCA 94  
|||||  
Db 112938 TCCTCTCTGGGATCCCA 112955  
RESULT 15  
AC007227/c  
LOCUS Homo sapiens chromosome 16 clone 54H19, WORKING DRAFT SEQUENCE, 17  
DEFINITION  
AC007227 177294 bp DNA HTG 06-APR-1999  
AC007227  
AC007227  
AC007227.1 GI:4567161  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 177294)  
Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,  
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,  
Han,C. and Deaven,L.  
Han,C. and Deaven,L.  
Sequencing of Human Chromosome 16p13.3  
Unpublished  
2 (bases 1 to 177294)  
Ricke,D.O.  
Large Scale Sequence Analysis and Annotation with the Sequence  
Comparison Analysis (SCAN) System  
Unpublished  
3 (bases 1 to 177294)  
Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,  
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,  
McMurry,K., Han,C. and Deaven,L.  
Direct Submission  
Submitted (06-APR-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los  
Alamos, NM 87545, USA  
COMMENT  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
690: contig of 690 bp in length  
gap of unknown length  
691 1381: contig of 691 bp in length  
gap of unknown length  
1382 2511: contig of 1130 bp in length  
gap of unknown length  
2512 3457: contig of 946 bp in length  
gap of unknown length  
3458 4261: contig of 804 bp in length  
gap of unknown length  
4262 5455: contig of 1194 bp in length  
gap of unknown length  
5456 6492: contig of 1037 bp in length  
gap of unknown length  
6493 6917: contig of 425 bp in length  
gap of unknown length  
6918 7518: contig of 601 bp in length  
gap of unknown length  
7519 7641: contig of 123 bp in length  
gap of unknown length  
7642 8312: contig of 671 bp in length  
gap of unknown length  
8313 9012: contig of 700 bp in length  
gap of unknown length  
9013 9661: contig of 649 bp in length  
gap of unknown length  
9662 12503: contig of 2842 bp in length  
gap of unknown length  
12504 49377: contig of 36874 bp in length  
gap of unknown length  
49378 97428: contig of 48051 bp in length  
gap of unknown length  
97429 177294: contig of 79866 bp in length.  
Location/Qualifiers  
1..177294  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="54H19"  
/chromosome="16"  
BASE COUNT 42026 a 44693 c 44677 g 45876 t 22 others  
ORIGIN  
Query Match 7.9%; Score 18; DB 35; Length 177294;  
Best Local Similarity 100.0%; Pred.No. 4.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 76 ATCCTCTCTGGGATCC 93  
|||||  
Db 164209 ATCCTCTCTGGGATCC 164192

Search completed: August 6, 1999, 11:39:19  
JOB time: 6445 sec



OM nucleic - nucleic search, using sw model

Title: US-09-049-696-11

Sequence: 1 GGCTTGTAGTGACAAAAA...AACAGGACACCAGCAAT 191

Scoring table: IDENTITY\_NUC

Database : N\_Geneseq\_36: \*

**Pred. NO.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

[illegible]

Sequence 10380 BP: 2210 A: 2953 C: 2851 G: 2366 T:

Query Match 15.9%; Score 30.4; DB 1; Length 10380;

Best Local Similarity 54.5%; Pred. No. 0.83; Mismatches 51; Indels 0; Gaps 0;

51 CATTGCTAAGCTTGGCAGCTGGAATACAGTCTGCAGACGACCTGCAACCTTGACCT 110  
DB 5823 CATCATGGGGCGGACGACTTCAATGATGACGACCCCTTCAGACCCCTCTACCT 5882  
111 GACTGTACGTCGCGCGTCCATGCTACCTGCTCCATATACAGTACT 162  
5883 TGCCGACGACGACCTGCGCTGATGACGACGACGATGACGATGACCT 5934

## RESULT 2

NS0114  
ID NS0114 standard; DNA: 2721 BP.

AC NS0114: 17-OCT-1991 (first entry)

DE DNA sequence encoding Epstein-Barr virus (EBV) outer surface protein.

KW Epstein-Barr virus; antigen; vaccine; ss.

OS Epstein-Barr virus.

FH Key Location/Qualifiers

FT mat\_peptide 1..2721

PN EP-151079-A.

PD 07-AUG-1985.

PF 28-JAN-1985; 400141.

PR 30-JAN-1984; US-575352.

PR 23-JUL-1984; US-633558.

PA (UICR) UNIV OF CHICAGO.

PI Klett E, Tanner J, Hummel M, Belsel C;

DR P-PSDB; P50073.

PT New fragment of Epstein-Barr Virus DNA - useful in vector to

PT express polypeptide for use in prepn. of vaccine against the

PT virus and for use in diagnosis.

PS Claim 1: Page 21-23; 26pp; English.

CC The sequence encodes an outer surface viral protein of EBV, used

CC to generate antibodies reacting with the surface proteins of

CC EBV-infected cells, and in the preparation of a vaccine against EBV.

CC Sequence 2721 BP; 762 A; 876 C; 557 G; 526 T;

Query Match 14.8%; Score 28.2; DB 1; Length 2721;

Best Local Similarity 59.3%; Pred. No. 2.9; Mismatches 33; Indels 0; Gaps 0;

106 ACCCTGACTGTACGTCGCGTCCATGCTACCTGCTCCATATACAGTACTTCC 165  
DB 1570 ACCCCAGCGAGTACTACACCCCAATGCCACGACCCGACGACGATGACTACC 1629  
166 AAACGACAAAGACACGACG 186  
1630 CCAACCCCAATGCCACGACG 1650

## RESULT 3

T04821  
ID T04821 standard; cDNA: 5931 BP.

DT 18-JAN-1996 (first entry)

DE EBV gp350/220 cDNA.

KW EBV; gp350; gp220; gp350/gp220; non-splicing variant; vaccine; ds.

OS Epstein-Barr virus.

FH Key Location/Qualifiers

FT cds 1014..3737

FT signal\_peptide 1014..1067

FT mat\_peptide 1068..3734

FT misc\_feature 2514..2515

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

FT misc\_feature 3105..3106

```

FT      /*tag= d
PT      /note= "Non-essential region suitable for cloning"
PN      WO9603508-A1.
PD      08-FEB-1996.
PF      26-JUL-1995: AU0453.
PR      26-JUL-1994: AU-007101.
PA      (CSTR ) COMMONWEALTH SCI & IND RES ORG.
PI      Both GW, Boyle DB, Vriati S;
DR      WPI: 96-117053/12.
PT      Viral vector derived from ovine adenovirus (OAV287) - used for
PT      recombinant prodn. and delivery of non-adenoviral polypeptide(s) or
PS      RNA molecules in host animals, esp. sheep.
CC      Claim 1: Fig 1: 53pp: English.
CC      The sequence represents the entire genome of ovine adenovirus OAV287,
CC      a benign adenovirus isolated from sheep in Australia, and begins at
CC      base 1 of the left-hand inverted terminal repeat. The sequence has
CC      been determined by cloning BamHI restriction fragments representing
CC      the entire genome and sequencing nested deletion mutants. Reading
CC      frames have been tentatively identified, including ELA/B, E4 and E3.
CC      The sequence contains promoters, inverted terminal repeats, viral
CC      packaging signals and RNA processing signals. The entire genomic
CC      clone is present in plasmid pOAV100. It is possible to introduce
CC      recombinant DNA at non-essential sites, e.g. in the E3 region,
CC      including a rotavirus VP3C antigen, Trichostyrnylus colubiformis
CC      17-kDa antigen, Terebia ovis 45W antigen or Lucilia cuprina p95
CC      antigen gene, or DNA encoding antisense RNA or a ribozyme. The
CC      sequence may be used in construction of vectors to express
CC      heterologous DNA in a grazing animal, e.g. a sheep. Many host
CC      species have no pre-existing immunity to the virus, allowing
CC      efficient infection, gene delivery and expression, and the
CC      possibility of virus spread to a non-target host is minimized. The
CC      virus may be used as a vector for recombinant vaccine production or
CC      transgenic animal construction.
SQ      Sequence 29544 BP; 10167 A; 5365 C; 4562 G; 9450 T;

```

```

Query Match      14.7% Score 28; DB 1; Length 29544;
Best Local Similarity 58.3%; Pred. No. 7.9;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY      4 TTTGTAGTGACAAAACCAATGCGCTACCTCCATCCAGGATGCTTAAGTT 63
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB      27819 TTTTTCCTTTTATACACACGACGATTCAGGCAATTCAGTTATGTTAATT 27878

QY      64 GGCACCTTGAATACAGCTGCAA 87
      || | | | | | | | | | | | | | | | | | | | | | | | | |
DB      27879 TGCTACGAACACTGCAGCACCA 27902

RESULT 5
Q41225
AC      Q41225 standard; DNA; 2190 BP.
DT      02-SEP-1993 (first entry)
DE      Clone PC14 encoding major surface gp of rat P. carinii.
KM      Major surface glycoprotein; gp116; rat; Pneumocystis carinii;
KW      vaccine; HIV; human immunodeficiency virus; diagnostic; PCR; ss.
OS      Rat Pneumocystis carinii.
PN      US7958683-A.
PD      01-APR-1993.
PF      09-OCT-1992: 958683.
PR      09-OCT-1992: US-958683.
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI      Angus CW, Edman JC, Kovacs JA, Powell F;
DR      WPI: 93-159487/19.
PT      Major surface glyco-protein of Pneumocystis carinii, encoded by
PT      multi-copy gene family - used in a vaccine and diagnostic assay
PT      for P. carinii infection in AIDS patients
PS      Disclosure; Page 36; 50pp: English.
CC      Multiple clones were identified by immunoscreening a rat P. carinii
CC      cDNA library using rat serum generated against whole rat P. carinii.
CC      Clones reactive with polyclonal serum were evaluated to identify
CC      those encoding gp116. Three such clones (PC3, PC5 and PC14) were

```

```

CC      sequenced and contained ORFs encoding closely related but distinct
CC      proteins. Although none of the clones contained the complete
CC      coding sequence, overlapping regions allowed alignment of the three
CC      clones and generation of a putative composite sequence encoding a
CC      protein of ca. 122 kD. The gp116 of P. carinii can be used in a
CC      vaccine against infection in HIV-infected individuals and also as a
CC      diagnostic agent. Study of the expression of the protein should lead
CC      to understanding its role in the pathogenesis of P. carinii pneumonia
CC      and may lead to new strategies for control or prevention of infection.
CC      The DNA sequence may be used to make PCR primers for diagnostic use.
SQ      See also Q41223-39.
      Sequence 2190 BP; 874 A; 320 C; 473 G; 523 T;

```

```

Query Match      14.5% Score 27.6; DB 1; Length 2190;
Best Local Similarity 53.8%; Pred. No. 4.3;
Matches 57; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY      73 AATATAGTCTGACAGCAAGCTTCAACCTTACCTGACTGTACGTCCTGGTCC 132
      | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1687 ACAGAGACGCAAAAAGAAATCTAACCTACTAGCAGCACCACACTAGCACCAG 1746

QY      133 AATGCTACCTGCGCTCCCAATTACAGTACTTCCAAACGACAGG 178
      ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB      1747 ACTACTACCAAGCACTACGACGACAACTACTACTACCAAGCCGG 1792

```

```

RESULT 6
T29408/c
ID      T29408 standard; cDNA; 2619 BP.
AC      T29408:
DT      14-OCT-1996 (first entry)
DE      Human metabotropic glutamate receptor mglur2 cDNA.
KM      Metabotropic glutamate receptor; antipsychotic; analgesic;
KW      anxiolytic; antidepressant; anticonvulsant; ss.
OS      Homo sapiens.
PI      Key location/Qualifiers
FT      cds 1..2616
      /*tag= a
FT      EP-711832-A2.
PN      PD 15-MAY-1996.
PF      09-NOV-1995: 308033.
PR      14-NOV-1994: US-337797.
PA      (ELIL ) LILLY & CO ELI.
PI      Burnett JP, Wayne NG, Sharp RL, Snyder YM;
DR      WPI: 96-232097/24.
PT      P-PSDB: R95052.
DE      Isolated human metabo: tropic glutamate receptor - used to identify
PT      agents useful as, e.g. antipsychotic, analgesic, anxiolytic,
PS      antidepressant and anticonvulsant agents
PS      Claim 2: Page 27-31; 44pp: English.
CC      DNA encoding the hMGR was obtained from a human fetal brain cDNA
CC      library. This nucleic acid may be expressed recombinantly in a host
CC      cell harbouring an expression vector. The encoded protein and
CC      transfected cells can be used for evaluating the effectiveness of a
CC      test compound, for use in the treatment or prevention of conditions
CC      associated with an excess or deficiency of stimulation of the hMGR.
SQ      Sequence 2619 BP; 439 A; 840 C; 760 G; 580 T;

```

```

Query Match      14.5% Score 27.6; DB 1; Length 2619;
Best Local Similarity 52.6%; Pred. No. 4.6;
Matches 60; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY      75 ATACAGTCTGACAGCAAGCTCAAACTTGCAGCTGACGTCCTCCGTCGTC 134
      | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      2430 AAGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2371

QY      135 TGCTACCTGCGCTCCCAATTACAGTACTTCCAAACGACAGGACGACG 188
      ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB      2370 CTGTACCGGTAGTACTGAGGTGACATAGAAAGATGCGCAAGATGCGCA 2317

```

Sequence	749 BP	234 Aa	82 Cc	150 Gg	283 Tt	
CC	a more specific and sensitive antibody response, and diagnosis,					
CC	responsible for relapsing fever or yaws (borreliosis). It should provided					
CC	sensu lato, but is absent from other Borrelia species (e.g. those					
CC	is involved in virulence and is highly conserved within B. burgdorferi					
CC	infections. In standard immunoassays or amplification/hybridization					
CC	(II). (II) and antibodies raised against (II) are used to diagnose such					
CC	other spirochaetes. (I), (II) and transformed cells are useful in					
CC	to specific nucleic acid sequences but not to genomic DNA from most					
CC	rabbit polyclonal antibody raised against a 13 kDa polypeptide of					
CC	Borrelia burgdorferi B313, but not reactive with most proteins from					
CC	other spirochaetes; and/or (ii) hybridizes under stringent conditions					
CC	to specific nucleic acid sequences but not to genomic DNA from most					
CC	other spirochaetes. (I), (II) and transformed cells are useful in					
CC	vaccines to protect against infection by B. burgdorferi sensu lato.					
CC	(II). (II) and antibodies raised against (II) are used to diagnose such					
CC	infections. In standard immunoassays or amplification/hybridization					
CC	tests. (I) are also used to produce recombinant (II). The 13 kD protein					
CC	is involved in virulence and is highly conserved within B. burgdorferi					
CC	sensu lato, but is absent from other Borrelia species (e.g. those					
CC	responsible for relapsing fever or yaws (borreliosis). It should provided					
CC	a more specific and sensitive antibody response, and diagnosis,					
CC	compared with use of whole bacteria as antigen.					
CC	Sequence	749 BP	234 Aa	82 Cc	150 Gg	283 Tt
FT	Location/Qualifiers	192..725				
FT	/*tag=	a				
FT						
PN	MO9912960-A2.					
PN	18-MAR-1999.					
PE	04-SEP-1998; IB1424.					
PR	16-SEP-1997; US-059036.					
PR	10-SEP-1997; DK-001041.					
PA	(STMB-) STMBICOM AB.					
PI	Bergstroem S;					
DR	WPI: 99-215027/18.					
DR	P-PSDB: Y04280.					
PT	Nucleic acid from Borrelia burgdorferi encoding virulence associated					
PT	protein P13					
PS	Claim 1; Page 113-114; 118pp; English.					
CC	The present sequence encodes a Borrelia garinii 1990 protein. The					
CC	present invention describes an isolated nucleic acid (I) that: (1)					
CC	encodes a polypeptide fragment (II) immunologically reactive with					
CC	rabbit polyclonal antibody raised against a 13 kDa polypeptide of					
CC	Borrelia burgdorferi B313, but not reactive with most proteins from					
CC	other spirochaetes; and/or (ii) hybridizes under stringent conditions					
CC	to specific nucleic acid sequences but not to genomic DNA from most					
CC	other spirochaetes. (I), (II) and transformed cells are useful in					
CC	vaccines to protect against infection by B. burgdorferi sensu lato.					
CC	(II). (II) and antibodies raised against (II) are used to diagnose such					
CC	infections. In standard immunoassays or amplification/hybridization					
CC	tests. (I) are also used to produce recombinant (II). The 13 kD protein					
CC	is involved in virulence and is highly conserved within B. burgdorferi					
CC	sensu lato, but is absent from other Borrelia species (e.g. those					
CC	responsible for relapsing fever or yaws (borreliosis). It should provided					
CC	a more specific and sensitive antibody response, and diagnosis,					
CC	compared with use of whole bacteria as antigen.					
CC	Sequence	749 BP	234 Aa	82 Cc	150 Gg	283 Tt

	Query Match	14.5%;	Score 27.6;	DB 1;	Length 749;	
	Best Local Similarity	55.1%;	Pred No. 2.9;			
Matches	54; Conservative	55.1%	0; Mismatches	44;	Indels	0; Gaps
Oy	69 TTGGAATACAGCTCGCAAGAAGTCCAAACCTTGACCCTGACTGTGCAGTCGGCCTGC	128				
Dd	713 TTtGGAAGAAGCACGCCAACAAACAAGAGACGTAAGTTGGCCCATGTATATCAAACCTAGG	654				
Oy	129 GTCCAATGCTACCTGCCCTCCAAATTACAGTAGTACTTCCA	166				
Dd	653 CTCAAATCTCCCAAGCGCTATATTAAAGACTATTTTTCA	616				
RESULT	8					
XI3060/c						
ID	XI3060 standard; DNA; 32768 BP.					
AC	XI3060:					
DT	19-MAR-1999 (first entry)					
DE	Enterococcus faecalis genome contig SEQ ID NO:123.					
KW	Enterococcus faecalis; contig; detection; Enterococcal infection; vaccine; attenuation; computer readable medium; ds. Enterococcus faecalis.					
OS	WO9850555-A2.					
PN						
PD	12-NOV-1998.					
PF	04-MAY-1998; UO8985.					
PR	14-NOV-1997; US-066009.					
PR	06-MAY-1997; US-044031.					
PR	16-MAY-1997; US-046655.					
(HUMA-) HUMAN GENOME SCI INC.						
Batrash SC, Dillon PJ, Kunsch CA;						

DB NPI; 99-045177/04.  
PT Isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
PT Claim 1; Page 739-755; 2084pp. English.  
CS A computer readable medium has been developed which has recorded on it  
CS 962 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
SQ Sequence 32768 BP; 10327 A; 5713 C; 7180 G; 9541 T;  
  
Query Match 14.2%; Score 27.2; DB 1; Length 32768;  
Best Local Similarity 47.2%; Pred No. 15;  
Matches 83; Conservative 0; Mismatches 93; Indels 0; Gaps 0

0Y 13 GACAAAAGCACAATGGCTTACTTCATAATCCAGGCAATGCTTAAGATTGGCACTTG 72  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 28927 GAGATTAACAGAATTGAATCCCCGCCCAACACAAATCTTCCTAAGATTCTCGTTG 28868

[illegible]

RESULT	9
ID	Q51080 standard; DNA; 2889 BP.
AC	Q51080;
DT	19-APR-1994 (first entry)
DE	Plasmid PRAP501 which encodes a haemin-binding protein from ORF1.
KW	Haemophilus somnus; immunogenic; haemolysin; LppB; LppC; thromboembolic meningoenophalitis; septicaemia; arthritis;
KW	pneumonia; ds.
OS	Haemophilus somnus.
FT	Key
FT	Location/Qualifiers
FT	29..628
FT	cds
FT	/*tag= a
FT	/note= "ORF2"
FT	734..1245
FT	/*tag= b
FT	/note= "ORF5"
FT	1246..1458
FT	/*tag= c
FT	/note= "ORF3"
FT	1474..1683
FT	/*tag= d
FT	/note= "ORF8"
FT	1679..2212
FT	/*tag= e
FT	/note= "ORF1, the HMB coding region"
FT	2208..2654
FT	/*tag= f
FT	/note= "ORF4"
FT	2491..2781
FT	/*tag= g
FT	/note= "ORF6"
FT	2777..2889
FT	cds

```

FT      /*tag= h
FT      /note="ORF7"
PN      WO9321323-A.
PD      28-OCT-1993.
PE      05-APR-1993: CA0135.
PR      09-APR-1993: US-865050.
PR      04-JUN-1992: US-893424.
PR      04-JUN-1992: US-893426.
PR      29-MAR-1993: US-038287.
PR      29-MAR-1993: US-038288.
PR      29-MAR-1993: US-038719.
PA      (UYSA-) UNIT SASKATCHEWAN.
PI      Harland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
PI      Rioux C, Theisen M;
DR      WPI: 93-351733/44.
DR      P-PSDB: R42370-7.
PT      Haemophilus somnus immunogenic proteins used in vaccines -
PT      selected from haemin-binding protein, haemolysin, Lppb and Lppc,
PT      and corresp. DNA
PS      Claim 8; Fig 2; 119pp: English.
CC      A genomic cosmid library of Haemophilus somnus HS25 DNA was screened
CC      for clones capable of binding bovine haemin and having haemolytic
CC      activity. Positive clones were subcloned various times, resulting
CC      in plasmid pRA501, which binds haemin but is not haemolytic. The
CC      clone was sequenced and was found to contain several open reading
CC      frames, potentially encoding 8 proteins. The haemin binding protein
CC      (encoded by the hmb gene) was encoded by ORF1. The protein can be
CC      used in vaccines for preventing or treating H. somnus infections,
CC      which cause thromboembolic meningo-encephalitis, septicemia,
CC      arthritis and pneumonia in vertebrates.
CC      See also Q51081-6.
SQ      Sequence 2889 BP; 1094 A; 608 C; 549 G; 638 T;

Query Match      14.1%; Score 27; DB 1; Length 2889;
Best Local Similarity 60.0%; Pred. No. 7.6;
Matches 45; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY      7 GTAGTGACAAAACACCAAAATGGCTCACTCCAAATCCAGCAGTGTGTTGGC 66
DB      1336 GTATTGGACATCAACCCACATATGACCACTTACCAATCCTCGCATGTGACGTCGTT 1395
OY      67 ACTTGAATACACT 81
DB      1396 ACTTGCTCAATAGT 1410

RESULT 10
N91628/c
ID      N91628 standard; DNA; 1077 BP.
AC      N91628;
DT      24-DEC-1990 (first entry)
DE      Sequence of gamma-4 region of human immunoglobulin genes sharing
DE      sigma-gamma-3 and sigma-gamma-4 C-mu C-sigma intron sequence
KM      J region; ss.
OS      Homo sapiens.
FH      Key
FH      Location/Qualifiers
FT      misc_feature
FT      107..1056
FT      /*tag= a
FT      /note="Sigma-gamma-4"

PN      MO8903425-A.
PD      20-APR-1989.
PE      14-OCT-1988: J01044.
PR      4-OCT-1987: JP-259206.
PR      09-JUL-1988: JP-170041.
PA      (EDGC-) Education Found Fujita (MITK MITH).
PI      Kurosawa Y, Awaya A, Ishizuka Y;
DR      WPI: 89-130043/17.
PT      Recombinant DNA switch sequences for immunoglobulin genes -
PT      enable specific control of prodn. of specific classes of antibody
PS      : Fig 9 B; 79pp: Japanese.
CC      The new switch sequences are introns. The DNA sequences are obtd. from
CC      suitable human cell lines, e.g. marrow cells, and pref. ARA-10 or Daudi,

```

```

CC      using a probe for the Ig(H) region, and cloned using a vector such as a
CC      plasmid or (pref.) lambda-phage. The sequences allow the control of prodn.
CC      of specific subclasses of Ig.
SQ      Sequence 1077 BP; 237 A; 292 C; 337 G; 211 T;

Query Match      14.0%; Score 26.8; DB 1; Length 1077;
Best Local Similarity 55.3%; Pred. No. 6.3;
Matches 52; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

OY      29 TGGCCTACCTCCAAATCCAGCATTCCTAAGTTGGACTTGGAATACACTGTCAG 88
DB      395 TGGCCTACCTCCAGGAGGACTCATCTGAATCTGATTCGATTCGAAGACTGTTCCAAA 336
OY      89 CAAGCTCACAAACCTTGACCTGACTGTCACGTC 122
DB      335 CAAGGTACATTCCTGTGCTGGGGGTTAGGAC 302

RESULT 11
O28398
ID      O28398 standard; DNA; 5946 BP.
AC      O28398;
DT      15-FEB-1993 (first entry)
DE      Rat nestin gene.
KW      Intermediate filament; central nervous system; brain tumour;
KW      neurofilament; ss.
OS      Rattus norvegicus.
FH      Key
FH      Location/Qualifiers
FT      cds
FT      128..5545
FT      /*tag= a
FT      /product= nestin
FT      /note= "128 could be 161"
FT      1..913
FT      exon
FT      914..1038
FT      /*tag= b
FT      /number= 1
FT      1039..1110
FT      /*tag= d
FT      /number= 3
FT      1111..5946
FT      /*tag= e
FT      /number= 4
FT      5929..5929
FT      /*tag= f

PN      MO9214821-A.
PD      03-SEP-1992.
PE      21-FEB-1992: U01375.
PE      22-FEB-1991: US-660412.
PA      (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI      Lendahl U, McKay RDG;
PI      WPI: 92-316175/38.
DR      P-PSDB: R27204.
PT      Diagnosis of pre-disposition to brain tumours - using DNA
PT      encoding nestin which distinguishes neural multi-potential stem
PT      cells from neuronal, glial and muscle cells
PS      Claim 3; Fig 1; 63pp: English.
CC      A lambda gt11 cDNA expression library was constructed from poly(A)+
CC      RNA from CNS of day 15 rat embryos. The library was screened with
CC      Mab Rat 401. cDNA from the most immunopositive clone was used as
CC      a probe to isolate additional clones from a lambda gt10 library
CC      from rat E15 CNS. Hybridising clones were sequenced; the remainder
CC      of the nestin gene was obtained from genomic sequences. There are
CC      two potential initiation codons at positions 128 and 160. Sequence
CC      comparison of cDNA and genomic clones reveals three introns at
CC      positions 912, 1038 and 1111. The sequence has significant
CC      similarities to the five classes of intermediate filaments.
CC      The coding sequence can be used to develop probes for
CC      detecting predisposition to brain tumours; the expression of nestin
CC      distinguishes neural multi potential stem cells and brain tumour
CC      cells from the more differentiated neural cell types (e.g. neuronal,

```



RESULT 14  
X05230/c  
ID X05230 standard; DNA; 2689 BP.  
AC X05230;  
DT 22-APR-1999 (first entry)  
DE HIV-2 genomic DNA sequence.  
KW HIV-1; HIV-2; Human immunodeficiency virus; primer/probe set; HIV;  
KW detection; nucleic acid amplification; PCR; RT PCR; OH PCR; virus;  
KW oligonucleotide hybridisation PCR; ss.  
OS Human immunodeficiency virus type 2.  
PN W09858086-A2.  
PD 23-DEC-1998.  
PF 04-JUN-1998; U11652.  
PR 16-JUN-1997; US-876346.  
PA (ABBO) ABBOTT LAB.  
PI Abiravaya K, Esping CAC, Gorzowski JJ, Hoenle RJ,  
PI Kroegeer PE, Moore JJ;  
DR MPI: 99-095352/08  
PT Nucleic acid primers and probes for detecting HIV 1 and HIV 2 -  
PT using known nucleic acid amplification procedures especially  
PT oligonucleotide hybridisation PCR  
PS Disclosure: Page 27-28; 35pp; English.  
CC The invention provides primer/probe sets for detecting HIV-1 or HIV-2  
CC that comprise two primers and at least one probe. The primer/probe sets  
CC are useful to detect HIV (i.e. HIV-1 and HIV-2, either separately or  
CC simultaneously) in biological samples using known nucleic acid  
CC amplification procedures e.g. PCR, reverse transcriptase (RT) PCR and  
CC especially oligonucleotide hybridisation PCR (OH PCR); HIV-1 and HIV-2  
CC are different viruses each with several subtypes due to the highly  
CC mutable nature of the virus (attributed to the inefficiency with which  
CC it converts its genetic material (RNA) into DNA to allow it to insert its  
CC genetic information into the host and/or recombination of viral genomes  
CC from different HIV populations). The sequences are designed to detect all  
CC known HIV-1 or HIV-2 subtypes by OH PCR and allow detection of target  
CC sequences which are DNA or, sequences which are embedded within the HIV  
CC genome and are therefore RNA by RT-PCR. Sequences X05232 to X05233  
CC represent specifically claimed primer and probe sequences present in  
CC eight different primer/probe sets that are used for detecting HIV-1 and  
CC HIV-2. The present sequence represents HIV-1 genomic DNA sequence that  
CC can be used as a target sequence in the method of the invention.  
SQ Sequence 2689 BP; 1066 A; 487 C; 597 G; 539 T;

Query Match 13.9%; Score 26.6; DB 1; Length 2689;  
Best Local Similarity 53.3%; Pred. No. 10;  
Matches 56; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 49 GCATGCTAGTGTGGCAATTGGAATATACGCTGCAAGCAAGCTACAACTTGACC 108  
DB 1756 GCCATTGCAAGAGCTTCTAATCTGCTTGTGATGGATGCTTCTAATCTTACC 1697

OY 109 CTGACTGTACAGTCCGTCGCTCAATGCTACCTGCTCCCAAT 153  
DB 1696 CTGCTCTCCCTCTACTATCTATATATACCTGCTTCCCTCTTTT 1652

RESULT 15  
T61290  
ID T61290 standard; cDNA; 2530 BP.  
AC T61290;  
DT 03-JUN-1997 (first entry)  
DE Mouse beta-3 integrin cDNA.  
KW Beta-3 integrin; bone resorption; osteoporosis; osteoclast;  
KW signal transduction; vitronectin receptor; ss.  
OS Mus sp.  
FH Key  
FT cds Location/Qualifiers  
FT 164..2527  
FT /+tag- a  
PN W09708316-A1.  
PD 06-MAR-1997.  
PF 27-AUG-1996; U13805.  
PR 31-AUG-1995; US-003020.

PA (MERI) MERCK & CO INC.  
PI Duong LT, Nutt EM, Rodan GA;  
DR MPI: 97-179273/16.  
DR P-PSDB; W13573.  
PT Mouse full length or truncated beta-3 integrin sub-unit(s) and  
PT coding sequences - useful for screening for cpds. that inhibit bone  
PT resorption  
PS Claim 5: Fig 1A-1D; 30pp; English.  
CC A cDNA clone (T61290) codes for the full-length mouse vitronectin  
CC integrin receptor beta-3 subunit (W13573). It was isolated from a  
CC mouse osteoclast cDNA library. Clones were obtd. by screening the  
CC library with a probe obtd. using primers (see also T61292-93) based  
CC on a highly conserved sequence amongst all integrin beta subunits,  
CC and the 3' cDNA of the clone was obtd. by PCR (see also T61294-95).  
CC A cDNA clone (T61291) for a novel truncated, soluble beta-3  
CC integrin (W13574) was also isolated. These cDNA clones can be used  
CC to produce (truncated) beta-3 integrin in host cells for use in  
CC assays to identify novel cpds. which bind the integrin, esp. for  
CC inhibiting bone resorption e.g. in osteoporosis.  
SQ Sequence 2530 BP; 599 A; 679 C; 709 G; 543 T;

Query Match 13.8%; Score 26.4; DB 1; Length 2530;  
Best Local Similarity 54.0%; Pred. No. 12;  
Matches 54; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 68 CTGGAAATACAGTCTGCAGCAAGCTCACAACCTTGACCTGACTGTAGTCCCGTG 127  
DB 1009 CCTGGATGGAAGACTGCGAGCAATGCTGCCCCAATGATGGCGACTGCACATTTGGCAC 1068

OY 128 CGTCCATGCTACCTGCGCTCCCAATTAAGTGAAGTCCCA 167  
DB 1069 CGACAACCACTACTGCTGCTCCCACTACCAAGGACTACCA 1108

Search completed: August 6, 1999, 04:03:27  
Job time: 1559 sec

---

**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:32:46 ; Search time 3075.15 Seconds  
(without alignments)  
116.101 Million cell updates/sec

Title: US-09-049-696-4

Perfect score: 181  
Sequence: 1 CAAAAGATGCACATTCATTA.....ACAGCAAAATCAAAATGC 181

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	177.8	98.2	480	35	AA581198	AA581198 nd38c07.r
2	71.2	39.3	427	34	AA498535	AA498535 vha0f12.r
3	36	19.9	205	31	AA290063	AA290063 vbs0a10.r
4	35.8	19.8	309	45	AI351964	AI351964 qq2a12.x
5	35.4	19.6	288	33	AA450336	AA450336 zx04g12.s
6	35	19.3	659	43	AI172714	AI172714 EST191.Ma
7	34.8	19.2	598	34	AA521299	AA521299 aa79f05.s
8	34.8	19.2	364	41	C93791	C93791 C93791 Dict
9	34.4	19.0	150	30	AA254006	AA254006 va07e09.r
10	34.2	18.9	551	35	AA549871	AA549871 094dm3.gm
11	33.8	18.7	416	30	AA255804	AA255804 zr83h06.r
12	33.8	18.7	452	54	HSM011299	HSM011299 Homo sapi
13	33.6	18.6	380	49	AU057922	AU057922 AU057922
14	33.2	18.3	641	43	AU030715	AU030715 AU030715
15	33	18.2	702	41	AI009297	AI009297 EST203748
16	33	18.2	526	43	AU028294	AU028294 AU028294
17	32.8	18.1	263	29	AA180346	AA180346 zp15c07.s
18	32.8	18.1	309	30	AA220766	AA220766 mv66g11.r
19	32.6	18.0	463	30	AA209580	AA209580 mw75g08.r
20	32.6	18.0	337	39	AA825396	AA825396 oc57g07.s
21	32.6	18.0	466	47	AU051371	AU051371 AU051371
22	32.6	18.0	481	48	AI580436	AI580436 tn79c08.x
23	32.4	17.9	221	30	AA254525	AA254525 val1a10.r
24	32.4	17.9	102	35	AA587120	AA587120 nt70b05.s
25	32.4	17.9	202	37	AA681489	AA681489 vr37h12.r
26	32.2	17.8	374	33	AA404760	AA404760 28158.Lam
27	32.2	17.8	256	48	AI590020	AI590020 tr74h10.x
28	32.2	17.8	431	50	AI667146	AI667146 fc38b11.x
29	32	17.7	456	22	HI5457	HI5457 ym29h03.s1
30	32	17.7	264	48	AI547520	AI547520 UI-R-C3-s
31	32	17.7	591	53	HSM002527	HSM002527 Homo sapi
32	31.8	17.6	352	33	AA390106	AA390106 vb28d08.r
33	31.8	17.6	438	37	WA3669	WA3669 23062.CD4-1
34	31.8	17.6	453	41	C94325	C94325 C94325 Dict
35	31.8	17.6	229	43	C97615	C97615 C97615 Dict
36	31.8	17.6	390	45	AI351737	AI351737 qt03e04.x
37	31.8	17.6	231	46	AI446139	AI446139 tj07d03.x
38	31.8	17.6	562	50	AI675898	AI675898 wc06h01.x
39	31.8	17.6	390	50	AI681318	AI681318 tx45h06.x
40	31.8	17.6	683	51	AI723590	AI723590 hcq1s165
41	31.6	17.5	401	33	AA400654	AA400654 zu70g10.s
42	31.6	17.5	495	33	AA400699	AA400699 zu70g10.s
43	31.6	17.5	379	36	AA610717	AA610717 np92c01.s
44	31.6	17.5	414	39	AA828323	AA828323 ob55f09.s
45	31.6	17.5	432	44	AI312542	AI312542 qp10c12.x

#### ALIGNMENTS

RESULT 1  
AA581198  
LOCUS  
DEFINITION  
nd38c07.r1 NCI\_CGAP\_Col Homo sapiens CDNA clone IMAGE:802572 5'  
similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.  
; mRNA sequence.  
AA581198  
ACCESSION  
ND 92358970

05-JAN-1998

VERSION	AA581198.1	GI:2358970
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 480)	
TITLE	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP)	
COMMENT	Tumor Gene Index	
	Unpublished (1997)	
	On May 9, 1995 this sequence version replaced gi:803001.	

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Michael  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies Inc., David Krizman,  
Ph.D.  
cDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LUNL at:  
[www-bio.linnl.gov/dbip/image/image.html](http://www-bio.linnl.gov/dbip/image/image.html)

Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 405.

**FEATURES**  
**Source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:802572"
/clone_1lb="NCI-CGAP_C01"
/tissue_type="bulk tumor"
/lab_host="DH10B"
/notes="Organ: colon; Vector: pCMV-Sport2; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dr."
BASE COUNT      168 a      101 c      97 g      114 t
ORIGIN

```

Query Match	98.2%	Score 177.8	DB 35	Length 480
Best Local Similarity	98.9%	Pred No. 1.9e-36		
Matches 179	Conservative	0	Mismatches 2	Indels 0
				Gaps 0

```
Oy      1 CAAAGATGCACATTCAATTAACTAACGAGACTCTATGAAAAAGGATGTGAGTTTGGTCT 60
        |||||
Db     102 CAAAGATGCACATTCAATAAAGTTACAGGACTCATGA AAAAGGATGTGAGTTTGGTCT 160
```

61 CCAATCCCGCAGACGAGAGGCTTATATGTTGCACAACATGTGATCTATAGT 120

Db 162 CCAATCCCGCCAGAGGAGAGGCTCTTAATGTTTGCAACAACAGTTGATTTATAGT 222

Qy 121 TGAATTCGTACAGAACAAACACAACAAGAGCTCCAAACAAGAAAATCAAAAATG 186

Db 222 TGAATCTGTACGAGACAAACCAACAAGAAGCTCCAAACAAGCAAAATCAAAAAATG 287

Qy	181 C 181
Db	282 C 282

## RESULT 2

AA498535	AA498535	427 bp	mRNA	EST	01-JUL-1997
LOCUS	AA498535				
DEFINITION	vh0f12.r1	Barstead mouse pooled organs	MPLRB4	Mus musculus	CCNA

clone IMAGE:889487 5' similar to TR:G1184066 G1184066  
CALCIUM-ACTIVATED CHLORIDE CHANNEL. ;, mRNA sequence

ACCESSION AA498535  
NID 92233558  
NID 92233558

VERSION AA498535.1 GI:2233558  
KEYWORDS EST.

SOURCE	ORGANISM
house mouse.	<i>Mus musculus</i>
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
	Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 427)
AUTHORS	Marta, M., Hiltner, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

The Waterbury-Watson ECM Product  
Theisting, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE	The Wasnu-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1294625

Contact: Alarra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: [mousteest@watson.wustl.edu](mailto:mousteest@watson.wustl.edu)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:517447

Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 377.

**FEATURES**  
**SOURCE**

```
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
```

```

/db_xref="taxon:10090"
/map="13"
/clone="IMAGE:889487"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"

```

/note-Orl; pooled). Vector: pT73D-Pac (Pharmacia) with a modified polylinker: site\_1: EcoRI; site\_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGACGATCTGTGAAGGGGAGCGCCGCCCTTTTTTTTTTTTTTTT TTTTTCAGTATCTGTGAAGGGGAGCGCGCCGCTTTTTTTTTTTTTTTT TTTT]; double-stranded cDNA was ligated to Eco RI adaptors [GTGGATTGCGGAC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bob Barstead."

BASE COUNT	143	a	108	c	83	g	93	t
ORIGIN								

Query Match	Score	DB	Length
39.3%	71.2	34	427
57.0%	40.0		

```

Best Local Similarity 75.9%; Pred No. 4e-09;
Matches 88; Conservative 0; Mismatches 28; Indels 0; Gaps 0

```

66 CCCGCCAGACGAGAGAGGCTTCATTAATGTTTGCACCAACATGTTGAATCTATAGTTGAAT 125  
 20 CACACCAAAAGAGAGGCTTCATCATGTTGTAAACCAAAATATCAATTCGTGGTTGAAT 79

**Q7** 126 TCCTACAGAACAAAACCACAAAGAAGCTCCAAACAGCAAAATCAAATAATGC 181  
||||| |||| | |||| | |||| | |||| | ||||  
**D8** TCCTACAGAAAAAATTCACATCAGAAGCCCCAAATGACCAAAACCAAGATGC 135

### RESULT 3

AA290063	AA290063	205 bp	mRNA	EST	14-APR-1997
LOCUS					
DEFINITION	vb50a10.r1 Soares mouse lymph node NBMLN Mus musculus cDNA clone				

IMAGE: 752346 5', mRNA sequence.  
ACCESSION AA290063  
1000000

NID	91936290	
VERSION	AA290063.1	GI:1936290
REVWORDS		
EST		

KEYWORDS	EST.
SOURCE	house mouse



TITLE Moore, B., Schellenberg, K., Seepoe, M., Tan, F.,  
 White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 Thaisang, B.  
 JOURNAL WashU-Merck EST Project 1997  
 COMMENT Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced g1:1407011.

**CONTACT:** Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [est@wustl.edu](mailto:est@wustl.edu)  
 This clone is available royalty-free through INL; contact the  
 IMAGE Consortium ([info@img.lnl.gov](mailto:info@img.lnl.gov)) for further information.  
 Seq primer: -4mln13 fwd. ET from Amersham  
 High quality sequence stop: 258.  
 Location/Qualifiers

```

source
1. .288
/organism="Homo sapiens"
/db_xref="GDB:5983160"
/db_xref="taxon:9606"
/map="92G05; 1; 1q32.1-1q32.3"
/clone_image:785542"
/clone_id="Soares_total_fetus_NB2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand
cDNA was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5']
TGTTTCAATCTGTAAGTGGAGCGCGCCCTTAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and W. Fatima Bonaldo. "

```

	Query Match	19.6%	Score 35.4;	DB 33;	Length 288;	
	Best Local Similarity	61.3%;	Pred. No. 6.3;			
	Matches 57; Conservative	0;	Mismatches 36;	Indels 0;	Gaps 0;	
Oy	86 TCTATATGTTGCACAAACATGTGATTCTANTAGTGAATTCGTACAAACAAAACCAC	145				
Db	99 TTGTGAAGCTTGACTGTACCATGTTTTCTGTTTAATTATGTAAAAACAACAACTAC	40				
Oy	146 AACAGAAGAGCTCCAAACACAAACCAATAATCAAAA	178				
Db	39 AACACACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	7				

RESULT		6				
Locus	A1172714/c					
DEFINITION	A1172714	659 bp	mRNA	EST	06-NOV-1998	
LOCUS	Esf19l Manduca sexta male antennae Uni-ZAP XR library Manduca sexta					
ACCESSION	CDDNA clone phsmac77_3'					
NID	A1172714					
VERSION	93f19569					
KEYWORDS	A1172714.1 GI:3719569					
SOURCE	EST.					
ORGANISM	tobacco hornworm.					
TITLE	Manduca sexta					
REFERENCE AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Endopterygota; Lepidoptera; Glossata; Diurysia; Sphingioidea; Spbingidae; sphinginae; Manduca.					
JOURNAL COMMENT	(bases 1 to 659) Robertson,H.M., Matros,R., Sears,C. and Nard,J.B. Expressed sequence tags from the antennae of male Manduca sexta moths reveal diversity of odorant binding proteins Unpublished (1998)					
	On Jan 14, 1998 this sequence version replaced gi 1798409.					

Other ESRs: ESR190  
Contact: Robertson HM  
Department of Entomology  
University of Illinois at Urbana-Champaign  
505 S. Goodwin, Urbana, IL 61801, USA  
Tel.: 217-243-0489  
Fax: 217-243-3499  
Email: hugh@debuc.edu  
Insert length: 950 Std Error: 0.00  
Seq primer: T7  
High quality sequence stop: 500.

FEATURES	SOURCE	Location/Qualifiers
		1. .659
		/organism="Manduca sexta"
		/db_xref="taxon:7130"
		/clone="pmsmac77"
		/clone_id="Manduca sexta male antennae Uni-ZAP XR library"
		/sex="male"
		/dev_stage="newly eclosed adults and pharate adults"
		/lab_host="X1 Blue MRF and SOLR"
		/note="Organ: antennae; Vector: Uni-ZAP XR; Site:1: ECORI; Site:2: XhoI; The library was prepared by Stratagene using oligo-T priming and unidirectional cloning with an adaptor at the 5' end (GGCAGCG) following the EcoRI site. The mRNA was prepared from antennae of late pupal and newly eclosed male moths. Clones were subcloned in vivo in mass into pBluescript maintained in SOLR cells for DNA sequencing."
BASE COUNT	192 a	252 t
ORIGIN	117 c	98 g

	Query Match	19.3%:	Score 35;	DB 43;	Length 659;
	Best Local Similarity	50.3%:	Pred. No. 7.3;		
Matches	86; Conservative	0;	Mismatches	85;	Indels 0; Gaps 0;
OY	8 TGCACATTCAMTAAAGTATACAGACTCATGTAAAAAGATGTAAGTTGTTCCCAATCC	67			
Dd	174 TGCTGTAGTGAGAAGGGCGCTCAAGATCATACTAATTATGATATATCCATGTGTTTAGT	115			
OY	68 CGCCAGACGGAGAMGGCTTTCTATTAATGTTTGACAACAATGTTGATTCATAGTGAATTC	127			
Dd	114 AACGAAACAAAAGGTGAAATTAGCATTTTTTTTATGTTATTTATGATTAATATAGATTTTGG	55			
OY	128 TGTATAGAACAAAACCACAACAAAGAACCTCCAACCAAGCAAATCAAAAA	178			
Dd	54 TTTCACAA	4			

AA521299/c RESULT 7  
LOCUS 598 bp mRNA EST 20-AUG-1997  
DEFINITION aa79f05..s1 NCI-CGAP-GCBI Homo sapiens cDNA clone IMAGE:827169 3'  
similar to contains Abi repetitive element; contains element MER22  
repetitive element ;, mRNA sequence.  
ACCESSION AA521299  
NID 92261842  
VERSION AA521299.1 GI:2261842  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 598)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
TITLE Unpublished (1997)  
JOURNAL On Sep 12, 1996 this sequence version replaced gi:1394702.  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550



```

/clone_lib="Soares mouse lymph node NbMLN"
/sex="male"
/tissue_type="lymph node"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
15',
TGTTACCAATCTGAAGTGGAGCGGCGCGGATTTTTTTTTTTTTTTTTTTT
3') double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M. Fatima Bonaldo."

```

	Query Match	Similarity	19.0%;	Score 34.4;	DB 30;	Length 150;
	Best Local	Similarity	57.4%;	Pred. No. 12;		
	Matches	62;	Conservative	0;	Mismatches 46;	Indels 0;
					Gaps	0;
OY	71	CAGACGGAGAGAGCTTCTATATGTTTGGACACAATGTTGATTCTATATGATTGATTCGT	130			
DB	33	CAGACAGGCTGTGACTTCTTACACACTTTCCCGAGGCTCTGATTAAGTCTTTAAATAAAA	92			
OY	131	ACAGACAACAAACCCACACAAAGAGCTCCAAACAGCAAAATATCAAAA	178			
DB	93	AA	140			

RESULT	10
LOCUS	AA549871
DEFINITION	0944m3 gmbpFHB3.1, G. Roman Reddy Plasmodium falciparum genomic clone 0944m, mRNA sequence.
ACCESSION	AA549871
NID	92320123
VERSION	AA549871.1
KEYWORDS	GI:2320123
SOURCE	EST.
ORGANISM	malaria parasite P. falciparum.
REFERENCE	Plasmodium falciparum.
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 551) Dane,J.B., Arnot,D.E., Bourke,P., Chakrabarti,D., Christodouloulova, Coppel,R., Comnan,A., Craig,A., Fischer,K., Foster,M., Goodman, Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A. Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su,X.-Z., Thompson,J.K., Vitali,F., Wellens,T.E. and Werner,E. Current status of the Plasmodium falciparum genome project Mol. Biochem. Parasitol. 79, 1-12 (1996)
JOURNAL	97001675
COMMENT	On May 18, 1995 this sequence version replaced gi:811199.

Contact: Dame JB  
 Department of Pathobiology  
 University of Florida  
 2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880  
 Tel: 352 392 4700  
 Fax: 352 392 9704  
 Email: dame@cdr.lfas.ufl.edu  
 Seq primer: T3.  
 Location/Qualifiers  
     1. 581  
     /organism="Plasmodium falciparum"  
     /db\_xref="taxon:5833"  
     /map="5"  
     /clone="0944m"  
     /clone\_lib="gmbpfn3.1, G. Roman Reddy"  
     /lab\_host="E. coli XL1-Blue"  
     /notes="vector: pBluescript SK(+); Genomic DNA, from  
     asynchronous blood stage parasites of the cloned Honduran

BASE COUNT	310 a	27 c	77 g	135 t	2 others
ORIGIN					
Query Match	18.9%	Score 34.2;	DB 35;	Length 551;	
Best Local Similarity	49.7%	Pred. No. 12;			
Matches	84;	Conservative	0;	Mismatches	85; Indels 0; Gaps 0.
QY	11 ACATTCATTAAGTACAGAGACTCTATGAAAAAGATGTGAGTTGTTCCATCCGCCG	70			
DB	330 ATAAATAAAAAGGAAGAACAATTCATCAATATATNGAAAAAGATATTGTGAAAAAATAAA	389			
QY	71 CAGAGCGAAGAGCGCTTCCTATATGTTTGCACACACATGTTGATTCTATATGTAATTCGT	130			
DB	330 GAAAGAGAAAGCATTTCTATAGAAAATTTATNGTATTTAAAAACATCTGAAAAACTAAAA	449			
QY	131 ACAGACAAAAACACACAAAGAGACCTCCAAACCAAGCAAAATCAAAAT	179			
DB	450 ATAGAAATTAAGAAAAAGAAAAAGAAATTTAGAAAATTTTAAATTAATAAAT	498			

RESULT	11
LOCUS	AA255804
DEFINITION	AA255806.r1 Soares.NHMPU_S1 Homo sapiens cDNA clone IMAGE:682331
ACCESSION	AA255804
NID	g1891345
VERSION	AA255804.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 416)
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S., Kicaula,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., Wille,Y., Wille,T., Waterston,R. and Wilson,R.
TITLE	Mash-Merck EST project 1997
COMMENT	Unpublished (1997) On Apr 14, 1993 this sequence version replaced g1:692984.

**CONTACT:** Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LIND; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Insert Length: 2440 Std Error: 0.00  
 Seq primer: -28m13 rev2 ER from Amersham.  
**FEATURES**  
 source  
 Location/Qualifiers  
 1..416  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="MMU16C3-C4 region; MMU16C3-C4 region"  
 /clone="IMAGE:682331"  
 /clone\_id="Scars\_NhMPu.S1"  
 /tissue\_type="Scars\_NhMPu.S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and  
 pregnant uterus"  
 /lab\_host="DH10B"  
 /note="Organ: mixed (see below); Vector: pRT73D-Pac  
 (Pharmacia) with a modified polylinker; Site\_1: Not I;







Mon Aug 9 13:26:17 1999

us-09-049-696-4.rst

---

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:02:11; Search time 921.96 Seconds  
(without alignments)  
49.118 Million cell updates/sec

Title: US-09-049-696-4

Perfect score: 181  
Sequence: 1 CAAAAGATGCACATTCATTA.....ACAGCAAAATCAAAATCC 181

Scoring table: IDENTITY\_NWC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	38.4	21.2	5156	1	V10238	Arabidopsis ferula
2	36.8	20.3	5156	1	T90505	Arabidopsis thalia
3	35	19.3	1794	1	T33873	Candida RAP1 CDNA
4	33.2	18.3	2312	1	V21646	T. gondii chorisma
5	32.8	18.1	286	1	X39448	Human secreted pro
6	32.6	18.0	1534	1	T15995	Human interleukin-
7	32.6	18.0	698	1	V06140	Viral infection ge
8	32	17.7	64976	1	V21209_16	Continuation (17 o
9	32	17.7	110000	1	X20248_00	Borrelia burgdorfe
10	31.2	17.2	1401	1	T72875	DNA encoding cyste
11	31.2	17.2	1401	1	V31108	Staphylococcus aur
12	31.2	17.2	6876	1	V74497	Staphylococcus aur
13	31	17.1	1838	1	T90504	Arabidopsis thalia
14	31	17.1	1838	1	V10237	Arabidopsis ferula
15	30.4	16.8	4004	1	O05680	DNA encoding B.t.
16	30.4	16.8	3414	1	O41745	btPS1245 gene. Ne
17	30.4	16.8	3414	1	O56922	Bacillus thuringie
18	30.2	16.7	707	1	X12975	Enterococcus faeca
19	29.6	16.4	1743	1	T85939	Shiella flexneri
20	29.6	16.4	2196	1	V59663	Human secreted pro
21	29.6	16.4	4137	1	X20278	Borrelia burgdorfe
22	29.2	16.1	7859	1	O12154	Shuttle vector pSW
23	29.2	16.1	7884	1	O25185	PSM6 expression ve
24	29.2	16.1	7859	1	O44265	PSM6 for expressio
25	29.2	16.1	1015	1	X00713	Human secreted pro
26	29.2	16.1	110000	1	X20248_01	Continuation (2 of
27	29	16.0	713	1	O48773	Aphrodite. DNA a
28	29	16.0	713	1	O49210	Hamster Aphrodite
29	29	16.0	1454	1	V23913	Plant CAD enzyme D
30	29	16.0	1474	1	V23914	Plant CAD enzyme D
31	29	16.0	620	1	V29636	Degenerate sequenc
32	29	16.0	2280	1	V52604	Human metalloprote
33	28.8	15.9	3414	1	O41744	DNA encoding B.t.
34	28.8	15.9	3414	1	O56921	Bacillus thuringie
35	28.8	15.9	1366	1	O44477	Pichia optimised B
36	28.8	15.9	2339	1	O63892	Tomato acid invert
37	28.8	15.9	2339	1	T01339	Tomato acid invert
38	28.8	15.9	5892	1	T79680	BRCA2 cancer suscep
39	28.8	15.9	7240	1	T79688	Partial BRCA2 cance
40	28.8	15.9	11283	1	T84841	Human breast and o
41	28.8	15.9	11385	1	T69707	Human breast cance
42	28.8	15.9	10485	1	X30256	Human BRCA2 (cm12)
43	28.8	15.9	10485	1	X30257	Human BRCA2 (cm13)

44 28.8 15.9 10485 1 X30258  
45 28.8 15.9 10485 1 X30259

Human BRCA2 (cm14)  
Human BRCA2 (cm15)

## ALIGNMENTS

RESULT	1	
ID	V10238/c	
AC	V10238	
DT	20-JUL-1998 (first entry)	
DE	Arabidopsis ferulate-5-hydroxylase genomic DNA.	
KW	Ferulate-5-hydroxylase; cinnamate-4-hydroxylase;	
KW	tissue-specific promoter; ligandin; syringyl; delignification;	
KW	pulping; transgenic plant; ss.	
OS	Arabidopsis thaliana.	
FH	Key	Location/Qualifiers
FT	CDS	2487..4522
FT		/tag= a
FT	Intron	/note= "contains introns"
FT		2997..3125
FT		/tag= b
FT		/number= 1
FT	Intron	3557..3901
FT		/tag= c
FT		/number= 2
PD	M09803535-A1.	
PN	29-JAN-1998.	
PF	18-JUL-1997: U12624.	
PR	16-DEC-1996: US-032908.	
PR	19-JUL-1996: US-022228.	
PA	(PURD ) PURDUE RES FOUND.	
PI	Chapple CCS.	
DR	WPI: 98-120689/11.	
P-PSDB:	M40099.	
PT	Tissue-specific regulatory plant promoter linked to	
PT	ferulate-5-hydroxylase gene - useful for manipulation of plant	
PT	ligandin composition	
PS	Claim 14; Page 48-51; 71pp: English.	
CS	This genomic DNA sequence includes a coding region, interrupted	
CC	by 2 introns, that codes for the ferulate-5-hydroxylase (F5H, see	
CC	M40099) of Arabidopsis thaliana. It comprises the HindIII-XhoI	
CC	fragment of pBIC20-F5H genomic clone isolated from a pBIC50	
CC	cosmid library using F5H cDNA (see V10237) as probe. F5H is a	
CC	cytochrome P450-dependent monooxygenase that catalyses the	
CC	meta-hydroxylation of ferulic acid in the monomer-specific pathway	
CC	branch required for sinapic acid and syringyl lignin biosynthesis.	
CC	A tissue-specific regulatory plant promoter, preferably derived	
CC	from the regulatory region of the C4H gene (see V10236) can be	
CC	linked to the F5H gene in order to increase the syringyl content of	
CC	ligandin in a transformed plant such as alfalfa, rice, maize, oilseed	
CC	rape, forage grasses, eucalyptus, pine, spruce, poplar, Arabidopsis	
CC	and tobacco. Increased levels of syringyl monomer residues in	
CC	ligandin renders the polymer more readily delignified so that the	
CC	plant is more readily pulped or digested.	
CC	Sequence. 5156 BP; 1671 A; 827 C; 906 G; 1752 T;	
SQ		
Query Match	21.2% Score 38.4; DB 1; Length 5156;	
Best Local Similarity	51.1% Pred. No. 0.033;	
Matches	90; Conservative 0; Mismatches 86; Indels 0; Gaps 0;	
OY	3 AAGATGCACATTCATTAAGTAACAGACATCATGAAAAAGATGATGTTCTCC 62	
DB	4835 AAAAAGTGAATAACTAAAAAATTTTTTTTTTTTAAAGTAAGAAAGTTCATCC 4776	
OY	63 AATCCCGCAGACGAGAGGCTTCTATATGTTTCACAACTGTGATTCATAGTG 122	
DB	4775 AAGATGATCTTGGACATGCTAAATTAATTAATTAAGAAAAAATCTATTATTACGTGA 4716	
OY	123 AATCTGTACAGAAACCAACCAAGAGGCTCCAAACCAAGCAAAATCAAAA 178	

Db 4715 CATTTCATTCAAGAAATCCCATTTAAAGAAAAAAGGCAACAAAGA 4660

## RESULT 2

190505/c  
ID T90505 standard; DNA: 5156 BP.  
AC T90505;  
DE 29-JAN-1998 (first entry)  
DE Arabidopsis thaliana ferulate-5-hydroxylase genomic DNA.  
DE Ferulate-5-hydroxylase: lignin: delignification: pulp: paper:  
KW feedstock; transgenic plant; alfalfa; Medicago; rice; Oryza;  
KW maize; Zea mays; oilseed rape; Brassica; forage grass; lath gene;  
KW tobacco; Nicotiana; eucalyptus; pine; Pinus; spruce; Picea;  
KW Poplar; Populus; crop improvement; Arabidopsis thaliana; ss.  
OS Arabidopsis thaliana ecotype Landsberg erecta.  
FH Key Location/Qualifiers  
FT CDS 2587..4522  
FT FT /\*tag= a  
FT FT /note= "includes introns"  
FT FT 2487..3002  
FT FT /\*tag= b  
FT FT /number= 1  
FT FT 3003..3130  
FT FT /\*tag= c  
FT FT /number= 1  
FT FT 3131..3557  
FT FT /\*tag= d  
FT FT /number= 2  
FT FT 3558..3902  
FT FT /\*tag= e  
FT FT /number= 2  
FT FT 3903..4522  
FT FT /\*tag= f  
FT FT /number= 3  
PM MO9723599-A2.  
PD 03-JUL-1997.  
PF 19-DEC-1996; U20094.  
PR 14-MAR-1996; US-013388.  
PR 22-DEC-1995; US-009119.  
PA (DUPO) DU PONT DE MEMOURS & CO E. I.  
PA (PUSD) PURDUE RES FOUN.  
PI Chapple C, C.  
PI WPI: 97-351042/32.  
DR P-PSDB: W26640.  
PT Nucleic acid encoding plant ferulate-5-hydroxylase - useful for  
PT altering guaiacyl:syringyl lignin monomer ratio in plants, to  
PT increase chemical delignification in paper and pulp making or to  
PT increase digestibility of feedstocks  
PS Claim 2; Page 29-31; 43pp; English.  
CC This genomic DNA sequence includes exons that encode the ferulate-  
CC 5-hydroxylase (F5H) (see W26640) of Arabidopsis thaliana. It  
CC comprises a HindIII-XhoI fragment of genomic clone pBIC20-F5H,  
CC which was isolated from a genomic library generated in binary  
CC cosmid vector pBIC20 by screening with a F5H cDNA insert (see  
CC T90504). A claimed chimeric gene that causes an altered  
CC guaiacyl:syringyl lignin monomer ratio in a transformed plant  
CC comprises the F5H cDNA or genomic clone linked either in the sense  
CC or antisense direction to at least one regulatory sequence. The  
CC gene allows the alteration of a plant's lignin content, especially  
CC to increase the syringyl lignin content. This renders the lignin  
CC more susceptible to chemical delignification of use to the paper  
CC and pulp industries. Alteration of the lignin content of grasses  
CC increases the digestibility for animal feedstocks. Suitable host  
CC plants include alfalfa, rice, maize, oilseed rape, forage grasses,  
CC tobacco, and tree crops such as eucalyptus, pine, spruce and poplar.  
SQ Sequence 5156 BP; 1670 A; 827 C; 907 G; 1752 T;

Query Match 20.3%; Score 36.8; DB 1; Length 5156;  
Best Local Similarity 50.6%; Pred. No. 0.094;  
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
QY AAAAGATGCATTCATTAAGTACAGAGACTATATGAAAAAGATGTGATTGTCTCC 62

Db 4835 AAAAGGGAATATACATTAAGAAAAATTTTTTTTTTTAAGTAGAAGTTCATCC 4776

QY 63 AATCCCGCAGCAGGAGAGAGCTCTTAATGTTGCACACAGTGTATCTATAGTTG 122

Db 4775 AAGATCGATCTTTGCACACATCTTAATTAATCTATGAAAAAAATCTTATTTTCTGTTA 4716

QY 123 AATCTGTACAGAACAAACCAACAAAGAGCTCCAAACGCAAAATCAAAA 178

Db 4715 CATTTCATTCAAGAAATCCCATTTAAAGAAAAAAGGCAACAAAGA 4660

## RESULT 3

190505/c  
ID T33873 standard; cDNA: 1794 BP.  
AC T33873;  
DE 24-NOV-1996 (first entry)  
DE Candida RAP1 cDNA.  
DE RAP1: rapamycin binding protein; FRP; immunosuppressive; agonist;  
KW fungicide; anti-mycotic; antagonist; cell proliferation; ss; ds.  
OS Candida albicans.  
FH Key Location/Qualifiers  
FT cds 1..1689  
FT FT /\*tag= a  
FT FT W09533052-A1.  
FT FT 07-DEC-1995.  
FT FT 30-MAY-1995; U06722.  
FT FT 27-MAY-1994; US-250795.  
FT FT 20-DEC-1994; US-360144.  
FT FT (MITO-) MITONIX INC.  
FT FT Berlin V, Chiu MI, Cottarel G, Damagnez V;  
FT WPI: 96-030563/03.  
DR P-PSDB: W04236.  
PT Rapamycin binding protein RAP1 - used as agonist, or antagonist of  
PT rapamycin cellular proliferation regulation  
PS Example 11; Page 92-95; 121pp; English.  
CC A cDNA clone (T33873) codes for a partial sequence (W04236) of a  
CC RAP1-like protein of Candida albicans. It was isolated from C.  
CC albicans cDNA in lambda Zap by PCR amplification using primers  
CC (T33887-88) based on conserved regions of RAP1 and TOR proteins  
CC (see also W04234-35 and W04237-38) and use of the PCR product to  
CC probe C. albicans cDNA libraries. The cDNA can be used for the  
CC prodn. of recombinant RAP1-like protein, esp. its rapamycin  
CC binding domain, or to design probes/primers used to detect  
CC Candida infection.  
SQ Sequence 1794 BP; 585 A; 273 C; 418 G; 518 T;

Query Match 19.3%; Score 35; DB 1; Length 1794;

Best Local Similarity 54.2%; Pred. No. 0.24;  
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 47 TGTGAGTTTGTCTCCAAATCCCGCAGAGGAGGCTTCTAATGTTTGCACACAT 106

Db 1372 TGGGGGTTTGAATTTCCCAAGGCGTTGGCTGATCAAGGGATATACGTTCCACAA 1431

QY 107 GTTGATTTATATGTTGAATTTCTGACAGAACAAACCAAGAGCTCCAAACAG 166

Db 1432 GTCAACACTGACGAAATTTTACGCGAGACAGATTGACGAAAAAGAGCTGAAGATTG 1491

QY 167 CAATCAAAA 177

Db 1492 CAAAAGCAAAA 1502

## RESULT 4

V21646  
ID V21646 standard; cDNA: 2312 BP.  
AC V21646;  
DE 21-JUL-1998 (first entry)  
DE T. gondii choriolate synthase enzyme encoding cDNA.  
KW Choriolate synthase; Apicomplexan parasite; poba gene; Toxoplasma gondii;  
KW metabolic pathway; parasitic disease; poba gene; Toxoplasma gondii;

```

PE 31-JUL-1998; IB1235.
PR 01-AUG-1997; US-905133.
PI (BEST ) GENSET.
PI Ductect A, Dumas Maline Edwards J, Iactrolx B;
DR WPI: 99-153781/13.
DR P-PSDB: Y11382.
PT New nucleic acids encoding human secreted - proteins obtained from
PT cDNA libraries prepared from substantia nigra, cerebellum, surrenals
PT and fetal brain tissue
PS Claim 1; Page 175; 434pp; English.
CC X39440 to X39597 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y1374 to Y1531,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 286 BP; 79 A; 45 C; 54 G; 103 T;

Query Match 18.1%; Score 32.8; DB 1; Length 286;
Best Local Similarity 49.4%; Pred. No. 0.63; Mismatches 81; Gaps 0;
Matches 81; Conservative 2; Indels 81;

QY 17 AATAAAGTAAAGAGAGACTGTATGAAAAAGAGTGGAGTTGTTCCATCCGCCAGAG 76
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 248 AAAATACACACACAGAGCTTTGGAGAGTGAATAGAGTTTCACTTCTGTTCCACATG 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 77 GAGAGAGCTTCTATATAGTTTGACCAACATGTTGTATGTTAGTTGAATCTGTACAGAA 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 188 GAATTGGCAAGATTCAATTGCGATTTTAGCTTTGACCAACAAATACAAACATCTMBVTYMAA 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 CAAACCCACAAAGAAAGCTCCAAACAGACAGAAATCAAAATG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128 CAGTTTCACTTCAATCAACATCAAGGCCAAGAGCAGATMAATG 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
T15995 T15995 standard; cDNA; 1534 BP.
AC T15995;
DT 24-JUL-1997 (revised)
DT 08-JUN-1996 (first entry)
DE Human interleukin-15 receptor cDNA.
KW Interleukin-15 receptor; IL-15R; murine; human; graft rejection;
KW treatment; prevention; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..756
FT FT /*tag= a
FT FT /product= immature_interleukin-15_receptor
FT FT 757..1534
FT FT /*tag= b
FT FT signal_peptide 1..33
FT FT /*tag= c
FT FT mat_peptide 34..753
FT FT /*tag= d
PN MO9530695-A.
PD 16-NOV-1995.
PE 04-MAY-1995; U05585.
PR 06-MAY-1994; US-236919.
PR 06-SEP-1994; US-300903.
RA (IMVP ) IMMUNEX CORP.

```

PI Anderson DM, Giri JG;  
DR WPI: 96-077255/08.  
P-PSDB: R91594.  
PT DNA encoding an interleukin-15 receptor - used to ameliorate undesired  
PT effects of interleukin-15, e.g. to prevent or treat graft rejection  
PS Claim 1; Page 23-34; 53pp; English.  
CC The cDNA shows a human open reading frame (ORF) encoding interleukin-15  
CC receptor (IL-15R) (see R91594), isolated from clone W5. It was isolated  
CC by cross-species hybridisation with the murine IL-15R cDNA (R15757). The  
CC mammalian IL-15R is encoded by a multi-exon gene. IL-15R variants can be  
CC attributed to different mRNA splicing events following transcription or  
CC from proteolytic cleavage of the IL-15R protein, where the IL-15R binding  
CC property is retained. The IL-15R protein is useful in treatment of  
CC undesired effects of interleukin-15, e.g. to prevent or treat graft  
CC rejection.  
CC (Revised entry submitted to correct sequence analysis breakdown.)  
SQ Sequence 1534 BP; 425 A; 445 C; 354 G; 310 T;

Query Match 18.0%; Score 32.6; DB 1; Length 1534;  
Best Local Similarity 58.9%; Pred. No. 1.1;  
Matches 56; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 84 CTCTATATGTTTGCACACATGTTGATCTATGATTCGTGACAGACAAAC 143  
DB 1433 CTGTGATTAACAAATTAACATCTATTTTAAATGCAAAAAA 1492  
OY 144 ACACACAGAGCTCCACACAGCAAAATCAAAA 178  
DB 1493 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1527

RESULT 7  
V06140  
ID V06140 standard; DNA: 698 BP.

AC 05-MAY-1998 (first entry)  
DE Viral infection gene SEQ ID NO:60.  
KW Viral infection; tumour suppressor; cellular gene; rat; cancer;  
KW serum protein; inhibitor; malignant phenotype; HIV; influenza;  
KW hepatitis; retrovirus; immunodeficiency; ds.  
OS Rattus sp.  
PN W05739119-A1.  
PD 23-OCT-1997.  
PE 11-APR-1997; U06067.  
PF 15-APR-1996; US-015334.  
PA (UYVVA-) UNIV VANDERBILT.  
PI Dubois RN, Organ EL, Rubin DH;  
DR WPI: 97-526456/48.  
PT Genes involved in viral infection and tumour suppression - used to  
PT develop products for reducing or preventing viral infection or for  
PT suppressing tumours  
PS Claim 1; Page 73; 101pp; English.  
CC The present sequence represents a viral infection gene. The present  
CC invention describes nucleic acid sequences isolated from rat. The  
CC sequences of the invention comprise 70 viral infection (VI) genes and  
CC 8 tumour suppressor (TS) genes. Propagating cell cultures in the  
CC absence of the serum protein (SP) allows selective elimination of cells  
CC persistently infected with a virus from the cell culture. Inhibitors of  
CC the SP can be used for reducing or inhibiting a viral infection.  
CC Inhibitors of the TS gene products can be used to suppress a malignant  
CC phenotype (MP). The methods and inhibitors can be used with viruses  
CC such as HIV, influenza, hepatitis virus or animal retroviruses such as  
CC simian immunodeficiency virus, avian immunodeficiency virus, bovine  
CC immunodeficiency virus, feline immunodeficiency virus, equine infectious  
CC anemia virus, canine arthritis encephalitis virus or vena virus.  
CC Because the identified genes are non-essential to cell survival, the  
CC treatment methods can be used in subjects without serious detrimental  
CC effects to the subjects.  
SQ Sequence 698 BP; 185 A; 137 C; 165 G; 180 T;

Query Match 18.0%; Score 32.6; DB 1; Length 698;

Best Local Similarity 69.8%; Pred. No. 0.9;  
Matches 44; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 116 ATAGTGAATTCGTGACAGACAAACCAACAAAGAGCTCCAAACAAATCAA 175  
DB 13 AGAGATGATTTAAACAAACAAACAAACAAACAAATCAACAAACAAACAAA 72  
OY 176 AAA 178  
DB 73 AAA 75

RESULT 8  
V21209\_16  
Continuation (17 of 17) of V21209 from base 1600001 (Methanococcus jannaschii circula  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209  
WP Fragment Name Begin End  
WP V21209\_00 1 110000  
WP V21209\_01 100001 210000  
WP V21209\_02 200001 310000  
WP V21209\_03 300001 410000  
WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000  
WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000  
WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976

Query Match 17.7%; Score 32; DB 1; Length 64976;  
Best Local Similarity 50.7%; Pred. No. 4.1;  
Matches 77; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

OY 26 ACAAGACTCTATGAAAAAGATGAGTTGTTCTCCATCCGCCAGACGAGAAAGCT 85  
DB 14577 AGAGTTTGTGAGATGAGAAATCTGTTATGTCACAAAGAAAACTATGAGTTGAT 14636  
OY 86 TCTATATGTTTGACACACATGTTGATCTATGATTAATTCGTACAGAAACAAACAC 145  
DB 14637 TTTTATATTTGGAGAAATTAAGAGAGTTATTCAGATGATGATTAATTAATGAGACC 14696  
OY 146 ACAAGAGCTCCAAACAAAGCAAAATCAAAA 177  
DB 14697 AATAAGAGAGAAATCAAAAGCTTATGAAA 14728

RESULT 9  
X20248\_00  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715  
ID X20248 standard; DNA: 910715 BP.  
AC X20248;  
DT 04-MAY-1999 (first entry)  
DE Borrelia burgdorferi; polynucleotide sequence #1.  
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;









---

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:40:45 ; Search time 650.92 Seconds  
(without alignments)  
73.414 Million cell updates/sec

Title: US-09-049-696-11

Perfect score: 191  
Sequence: 1 GCGCTGTGATGACACAAAAA.....AACAGACACACCAATT 191

Scoring table: OLIGO-MWC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	9.4	2040	1 T43949	Sequence flanking
2	17	8.9	325	1 V78785	Staphylococcus aur
3	17	8.9	305	1 V78656	Staphylococcus aur
4	17	8.9	14078	1 V74502	Staphylococcus aur
5	17	8.9	357	1 V78411	Staphylococcus aur
6	16	8.4	5187	1 T14218	Mouse patched gene
7	16	8.4	398	1 T21837	Human gene signatu
8	16	8.4	2237	1 T38267	200 gene different
9	16	8.4	5187	1 V21589	Mouse patched (pic
10	16	8.4	5187	1 V21587	Precis coenila (but
11	16	8.4	5187	1 V64092	Mouse patched gene
12	16	8.4	5187	1 V64099	Precis coenila patc
13	16	8.4	529	1 V87937	EST clone FC708. N
14	16	8.4	438	1 X41096	Human secreted pro
15	15	7.9	1398	1 N91208	Acetyl-CoA reducta
16	15	7.9	134525	1 O41525	Total base sequenc
17	15	7.9	2094	1 O10500	Acetoacetyl CoA re
18	15	7.9	1805	1 O14956	Cytochrome P 450 C
19	15	7.9	775	1 M40220	DNA sequence inclu
20	15	7.9	5270	1 Q25387	Rat thyrotropin re
21	15	7.9	1129	1 Q48616	Cpg depleted laci
22	15	7.9	2401	1 Q48539	Bacillus subtilis
23	15	7.9	110000	1 T58840-1	Continuation (2 of
24	15	7.9	72	1 T59111	DNA polymerase 11g
25	15	7.9	2803	1 T62384	Human A33 antigen
26	15	7.9	2565	1 T62369	Human A33 antigen
27	15	7.9	2044	1 T75623	Canine herpesvirus
28	15	7.9	3374	1 T20872	Murine phospholipa
29	15	7.9	2799	1 V20871	Murine phospholipa
30	15	7.9	1448	1 T84017	DNA encoding a sul
31	15	7.9	1448	1 V53413	DNA encoding a sir
32	15	7.9	116624	1 V52850	Human eyal gene co
33	15	7.9	6703	1 V49536	Adenylate cyclase
34	15	7.9	5240	1 V48268	P. chrysogenum bet
35	15	7.9	2044	1 V66945	Canine herpes viru
36	15	7.9	2044	1 V66946	Canine herpes viru
37	15	7.9	420	1 V66947	Canine herpes viru
38	15	7.9	1312	1 V63102	Crepis sp. delta-1
39	15	7.9	550	1 V63103	Vernonia galamensi
40	15	7.9	597	1 V89069	EST clone B261. Ne
41	15	7.9	357	1 V88194	EST clone FX65. Ne
42	15	7.9	13508	1 V74431	Staphylococcus aur
43	15	7.9	32768	1 X13336	Enterococcus faeca

## ALIGNMENTS

RESULT 1	1	44	15	7.9	17082	1	X13166
T43949		45		7.9	121	1	X12793
AC	T43949						Enterococcus faeca
AC	T43949						Human diallelic po
DE	18-AUG-1997 (first entry)						
DE	Sequence flanking marker 63-2 in HH region of chromosome 6p2.1.						
KW	Primer; polymerase chain reaction; amplifi; hereditary haemochromatosis;						
KW	HH; mutation; HH-associated allele; base-pair polymorphism; HHP-1;						
KW	HHP-19; HHP-29; microsatellite repeat allele; genetic marker;						
KW	interferon treatment; hepatitis C infection; ss.						
OS	Synthetic.						
PN	MO635803-AL.						
PN	14-NOV-1996.						
PF	08-MAY-1996; U06583.						
PR	08-MAY-1995; US-436074.						
PR	15-NOV-1995; US-559302.						
PR	09-FEB-1996; US-599252.						
PA	(MERC-) MERCATOR GENETICS INC.						
PI	Drayna DT, Feder JN, Gaithe A, Kimmel BE, Thomas WJ;						
PI	Wolff RK;						
DR	WPI: 96-518691/51.						
PT	Diagnosing and genotyping of hereditary haemochromatosis (HH) -						
PT	using primers to detect specific polymorphisms of the HH gene on						
PT	chromosome 6p2.1 or novel microsatellite markers						
PS	Disclosure: Fig 1Y; 67pp; English.						
CC	The sequences given in T43925-55 represent portions of the genome						
CC	surrounding several markers of the invention. The markers were						
CC	identified using the series of primer pairs given in T71901-72						
CC	which were used to determine the presence or absence of the common						
CC	hereditary haemochromatosis (HH) gene mutation in an individual. The						
CC	method comprised assessing genomic DNA from an individual for the						
CC	presence or absence of the HH-associated allele of the single base-pair						
CC	polymorphism HHP-1, HHP-19 or HHP-29, and/or at least one non-optimal						
CC	marker comprising the following microsatellite repeat alleles of group						
CC	A and optionally of group B:						
CC	Group A: 19D9, 18B4, 1A2, 1E4, 2A2E, 2B8, 3321-1, 4073-1, 4440-1, 4440-2,						
CC	731-1, 5091-1, 3216-1, 4072-2, 950-1, 950-2, 950-3, 950-4, 950-5, 950-6,						
CC	950-8, 63-3, 63-2, 63-3, 65-1, 65-2, 373-8, 373-29, 66-1, 241-6, 241-29;						
CC	Group B: D6S464, D6S306, D6S258, D6S265, D6S105 and D6S1001.						
CC	The absence of the genotype indicates the likelihood of the presence of						
CC	the HH mutation. Knowledge of the new genetic markers allows the						
CC	definition of genotypes characteristic of heterozygous carriers and						
CC	homozygotes having a HH mutation in their genomic DNA. The potential for						
CC	HH in an individual interferes with the effectiveness of interferon						
CC	treatment for hepatitis C infection. By diagnosing this potential, the						
CC	responsiveness of interferon treatment may be evaluated.						
SO	Sequence 2040 BP; 713 A; 409 C; 398 G; 513 T;						
Query Match	9.4%; Score 18; DB 1; Length 2040;						
Best Local Similarity	100.0%; Pred. No. 0.79;						
Matches	18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	3 CTTGTGATGACACAAAA 20						
DB	195 CTTGTGATGACACAAAA 212						
RESULT 2							
V78785							
AC	V78785 standard; DNA: 325 BP.						
AC	V78785; (first entry)						
DE	Staphylococcus aureus contig SEQ ID #4474.						
KW	Computer readable medium; vaccine; S. aureus infection; immunodetection;						
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;						

KW skin infection; surgical wound infection; scalded skin syndrome;  
 KM toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PE 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 DR WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1: Page 2979; 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 SQ Sequence 325 BP; 103 A; 55 C; 62 G; 102 T;

Query Match 8.9%; Score 17; DB 1; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 2.7;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TTGTAGTGACAAAAC 21  
 ||||||||||||||||  
 DB 171 TTGTAGTGACAAAAC 187

RESULT 3  
 V78656/c  
 ID V78656 standard; DNA; 305 BP.  
 AC V78656;  
 DE 16-MAR-1999 (first entry)  
 KM Staphylococcus aureus contig SEQ ID #4345.  
 KM Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KM skin infection; surgical wound infection; scalded skin syndrome;  
 KM toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PE 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 DR WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1: Page 2927; 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 SQ Sequence 305 BP; 92 A; 48 C; 48 G; 114 T;

Query Match 8.9%; Score 17; DB 1; Length 305;  
 Best Local Similarity 100.0%; Pred. No. 2.6;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TTGTAGTGACAAAAC 21  
 ||||||||||||||||  
 DB 55 TTGTAGTGACAAAAC 39

RESULT 4  
 V74502/c  
 ID V74502 standard; DNA; 14078 BP.  
 AC V74502;  
 DE 16-MAR-1999 (first entry)  
 KM Staphylococcus aureus contig SEQ ID #191.  
 KM Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KM skin infection; surgical wound infection; scalded skin syndrome;  
 KM toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 FH Key  
 FT misc-feature  
 FT Location/Qualifiers  
 FT 1381..1440  
 FT /tag- a  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT 3181..3240  
 FT /tag- b  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 4981..5040  
 FT /tag- c  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 6781..6840  
 FT /tag- d  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 8581..8640  
 FT /tag- e  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"  
 FT misc-feature  
 FT 10381..10440  
 FT /tag- f  
 FT /note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

```

FT misc_feature 12181. 12240
FT /*tag- 9
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 13981. 14040
FT /*tag- h
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997.
PR 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA.
PI WPI; 97-374922/35.
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 887-895; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 14078 BP; 4438 A; 2440 C; 2355 G; 4363 T;

Query Match 8.9%; Score 17; DB 1; Length 14078;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 TTGTAGTGACAAAAC 21
DB 6739 TTGTAGTGACAAAAC 6723

RESULT 5
V78411/c
ID V78411 standard; DNA: 357 BP.
AC V78411;
DE 16-MAR-1999 (first entry)
DT Staphylococcus aureus contig SEO ID #4100.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
PN EP-786519-A2.
PD 30-JUL-1997.
PR 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA.
PI WPI; 97-374922/35.

```

```

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 2825; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 357 BP; 103 A; 69 C; 59 G; 123 T;

Query Match 8.9%; Score 17; DB 1; Length 357;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 TTGTAGTGACAAAAC 21
DB 215 TTGTAGTGACAAAAC 199.

RESULT 6
T14218/c
ID T14218 standard; cDNA: 5187 BP.
AC T14218;
DE 11-JUL-1996 (first entry)
DT Mouse patched gene.
KW Patched gene; plc gene; embryo development; cellular regulation;
KW signal transduction; ligand; antibody; hedgehog protein;
KW gene therapy; ss.
OS Mus musculus.
FH Key
FT cds 97.4401
FT /*tag- a
FT /*tag- b
FT /*tag- c
FT /note- "bases 4536 and 4537 are given as k in the
FT specification"
FT misc_difference 4563
FT /*tag- c
FT /note- "base 4563 is given as r in the
FT specification"
FT misc_difference 4563
FT /*tag- c
FT /note- "base 4563 is given as r in the
FT specification"
FN W09611260-A1.
PD 18-APR-1996.
PR 06-OCT-1995; U13233.
PR 07-OCT-1994; US-319745.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Goodrich LV, Johnson RL, Scott MP;
PI WPI; 96-209842/21.
DR P-PSDB; R94380.
DT DNA encoding patched protein other than Drosophila patched protein
PT used to produce antibodies which detect or inhibit patched protein
PT ligand signal transduction in cells
PS Disclosure; Page 43-46; 70pp; English.
CC A cDNA clone (T14218) coding for the mouse patched protein (PTC)
CC (R94380) was obtd. by PCR amplification of cDNA derived from
CC a mouse limb bud, and use of the PCR product to probe a mouse 8.5
CC dpc lambda-gt10 cDNA library. The PCR primers (see T14226-27)
CC were based on patched amino acid sequences from fly, mosquito (see
CC also R94384), butterfly (see also R75373) and beetle (see also

```

CC R94385). The mouse p1c gene can be used for prodn. of large amounts  
 CC of recombinant p1c, as a probe e.g. to detect gene mutations, in  
 CC gene therapy, to study embryo development, to produce transgenic  
 CC animal models, etc.  
 SQ Sequence 5187 BP: 1162 A; 1411 C; 1332 G; 1279 T;

Query Match 8.4%; Score 16; DB 1; Length 5187;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 CAAAGACCCAAATG 30  
 DB 5069 CAAAGACCCAAATG 5054

RESULT 7  
 ID T21837 standard; cDNA to mRNA; 398 BP.  
 AC T21837;  
 DE 01-AUG-1996 (first entry)  
 KW Human gene signature HUMG50378.  
 KW Gene signature: messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 OS Homo sapiens.  
 PN M09514772-A1.  
 PD 01-JUN-1995.  
 PF 11-NOV-1994; J01916.  
 PR 12-NOV-1993; JP-355504.  
 PA (MATS/) MATSUBARA K.  
 PI (OKUB/) OKUBO K.  
 PI Matsubara K, Okubo K;  
 DR WPI: 95-206931/27.  
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 PS Claim 1: Page 978; 2245pp; Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in T19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SQ Sequence 398 BP: 108 A; 103 C; 108 G; 77 T;

Query Match 8.4%; Score 16; DB 1; Length 398;  
 Best Local Similarity 100.0%; Pred. No. 9.5;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 173 ACAAGGACCCAGCA 188  
 DB 17 ACAAGGACCCAGCA 32

RESULT 8  
 ID T38267 standard; cDNA; 2237 BP.  
 AC T38267;  
 DE 29-DEC-1996 (first entry)  
 DE 200 gene differentially expressed in T helper cells.  
 KW T helper cell; TH cell; T-cell; T-lymphocyte; 200 gene;

KW differential expression; immune disorder; multiple sclerosis;  
 KW asthma; lepromatous leprosy; diagnosis; therapy; receptor; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 42..947  
 FT signal\_peptide 42..101  
 FT mat\_peptide 102..944  
 FT msc\_difference 1298  
 FT /tag= d  
 FT /note= "base 1298 is given as r in the  
 FT specification"  
 PN W09627603-A1.  
 PD 12-SEP-1996.  
 PF 01-MAR-1996; U02798.  
 PR 03-MAR-1995; US-398633.  
 PR 07-JUN-1995; US-487748.  
 PA (MIL-) MILLENNIUM PHARM INC.  
 PI Levinson DA;  
 DR WPI: 96-433404/43.  
 DR P-PSDB; W01049.  
 PT Genes and their products differentially expressed in T helper cells  
 PT - useful in diagnosis and treatment of immune disorders, e.g.  
 PT multiple sclerosis, asthma, lepromatous leprosy, etc.  
 PS Claim 1: Fig 24; 218pp; English.  
 CC Novel human gene 200 (T38267) was isolated from a lambda gtl1  
 CC human lymphocyte cDNA library using the murine 200 gene (see also  
 CC T38265) as probe. The gene codes for a novel cell surface  
 CC receptor (W01049) of the Ig superfamily class. The 200 gene is  
 CC expressed at levels many-fold higher in T helper Th1 cell  
 CC subpopulations than in Th2 cell subpopulations. The gene and its  
 CC product can be used therapeutically to ameliorate immune disorders  
 CC and in the diagnosis of such disorders. It can also be used to  
 CC produce soluble 200 gene-Ig fusions and to breed transgenic animals  
 CC useful as models of TH cell subpopulation-related disorders.  
 SQ Sequence 2237 BP: 587 A; 518 C; 525 G; 606 T;

Query Match 8.4%; Score 16; DB 1; Length 2237;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 TCCAAATCCAGCAT 53  
 DB 376 TCCAAATCCAGCAT 391

RESULT 9  
 ID V21589/c  
 AC V21589;  
 DE 25-JUN-1998 (first entry)  
 DE Mouse patched (ptc) protein encoding cDNA.  
 KW Patched protein; ptc; cancer; tumour suppressor; cell adhesion promoter;  
 KW wound healing; ageing; mouse; ss.  
 OS Mus sp.  
 FH Key Location/Qualifiers  
 FT CDS 97..4401  
 FT /tag= a  
 FT /product= "mouse patched protein"  
 PN W09745541-A2.  
 PD 04-DEC-1997.  
 PF 02-JUN-1997; U09553.  
 PR 31-MAY-1996; US-656055.  
 PA (REGC ) UNIV CALIFORNIA.  
 PA (STRD ) UNIV IELAND STANFORD JUNIOR.  
 PI Epstein E, Goodrich LV, Johnson RL, Oro A, Scott MP;  
 DR WPI: 98-032648/03.  
 DR P-PSDB; W52199.  
 PT Patched protein other than Drosophila melanogaster patched protein -  
 PT used for characterising the phenotype of a tumour

PS Claim 5: Pages 63-66: 86pp: English.  
 CC This genomic DNA encodes a mouse patched (ptc) protein. This can be used  
 CC to construct an expression cassette comprising an altered patch or  
 CC hedgehog gene. The expression cassette comprises a nucleic acid encoding  
 CC a patched protein other than a Drosophila melanogaster patched protein,  
 CC or fragment of at least 12 nucleotides in length, as other than an intact  
 CC chromosome under transcriptional control of a transcriptional initiation  
 CC region, and a transcriptional termination region, both functional in an  
 CC expression host. A genetically engineered mammalian cell comprising this  
 CC expression cassette as an extrachromosomal element or integrated into the  
 CC genome of the cell can be predisposed to develop basal cell carcinoma as  
 CC a result of the transfection. By analysing DNA, functional analysis of  
 CC patched protein function, or by detecting antibody binding to abnormal  
 CC patched protein, a genetic predisposition to developmental abnormalities  
 CC and cancer can be diagnosed. This analysis can also be used for  
 CC characterising the phenotype of a tumour, particularly a carcinoma,  
 CC especially a basal cell carcinoma. The methods can also be used for  
 CC characterising transitional cell carcinoma of the bladder, meningiomas  
 CC medulloblastomas, etc. The modified cells comprising the expression  
 CC cassette can be used to determine the role of different exons of the  
 CC patched gene in oncogenesis, signal transduction, etc. Transgenic animal  
 CC models created from these cells can be used as animal models for  
 CC carcinomas of the skin. The patched protein of mosquito, butterfly or  
 CC beetle or alternatively, a mammalian patched protein of human or mouse  
 CC can be used to identify ligands or substrates that bind to, modulate, or  
 CC mimic the action of patched gene. These agents could be used as tumour  
 CC suppressors, cell adhesion promoters (e.g. in wound healing and ageing).  
 SQ Sequence 5187 BP; 1165 A; 1412 C; 1338 G; 1270 T;

Query Match 8.4%; Score 16; DB 1; Length 5187;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 CAAAACACCAAAATG 30  
 ||||||||||||||||  
 DB 5069 CAAAACACCAAAATG 5054

RESULT 10  
 V21587/c  
 ID V21587 standard; DNA: 5187 BP.  
 AC V21587;  
 DT 25-JUN-1998 (first entry)  
 DE Precis coenia (butterfly) patched (ptc) protein encoding DNA.  
 KM Patched protein; ptc; cancer; tumour suppressor; cell adhesion promoter;  
 KM wound healing; ageing; Precis coenia; butterfly; ss.  
 OS Precis coenia.  
 PN W09745541.R2.  
 PD 04-DEC-1997.  
 PF 02-JUN-1997; U09553.  
 PR 31-MAY-1996; US-656055.  
 PA (REGC) UNIV CALIFORNIA.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Epstein E, Goodrich LV, Johnson RL, Oto A, Scott MP;  
 PI MPI: 98-032648/03.  
 DR Patched protein other than Drosophila melanogaster patched protein -  
 PT used for characterising the phenotype of a tumour  
 PS Claim 2: Pages 49-51: 86pp: English.  
 CC This DNA is stated to encode a Precis coenia patched (ptc) protein and  
 CC is identical to the mouse ptc cDNA. This can be used to construct an  
 CC expression cassette comprising an altered patch or hedgehog gene. The  
 CC expression cassette comprises a nucleic acid encoding a patched protein  
 CC least 12 nucleotides in length, as other than an intact chromosome under  
 CC transcriptional control of a transcriptional initiation region, and a  
 CC transcriptional termination region, both functional in an expression  
 CC host. A genetically engineered mammalian cell comprising this expression  
 CC cassette as an extrachromosomal element or integrated into the genome of  
 CC the cell can be predisposed to develop basal cell carcinoma as a result  
 CC of the transfection. By analysing DNA, functional analysis of patched  
 CC protein function, or by detecting antibody binding to abnormal patched  
 CC protein, a genetic predisposition to developmental abnormalities and

CC cancer can be diagnosed. This analysis can also be used for  
 CC characterising the phenotype of a tumour, particularly a carcinoma,  
 CC especially a basal cell carcinoma. The methods can also be used for  
 CC characterising transitional cell carcinoma of the bladder, meningiomas  
 CC medulloblastomas, etc. The modified cells comprising the expression  
 CC cassette can be used to determine the role of different exons of the  
 CC patched gene in oncogenesis, signal transduction, etc. Transgenic animal  
 CC models created from these cells can be used as animal models for  
 CC carcinomas of the skin. The patched protein of mosquito, butterfly or  
 CC beetle or alternatively, a mammalian patched protein of human or mouse  
 CC can be used to identify ligands or substrates that bind to, modulate, or  
 CC mimic the action of patched gene. These agents could be used as tumour  
 CC suppressors, cell adhesion promoters (e.g. in wound healing and ageing).  
 SQ Sequence 5187 BP; 1165 A; 1412 C; 1337 G; 1270 T;

Query Match 8.4%; Score 16; DB 1; Length 5187;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 CAAAACACCAAAATG 30  
 ||||||||||||||||  
 DB 5069 CAAAACACCAAAATG 5054

RESULT 11  
 V64092/c  
 ID V64092 standard; cDNA: 5187 BP.  
 AC V64092;  
 DT 26-JAN-1999 (first entry)  
 DE Mouse patched gene.  
 KM Mouse; patched gene; diagnosis; treatment; developmental disorder;  
 KM cancer; healing; injured tissue; ptc; spina bifida; Wnt-1 oncogene;  
 KM sperm production; gene therapy; ss.  
 OS Mus sp.  
 FH Key  
 FT CDS Location/Qualifiers  
 FT CDS 97..4401  
 FT /tag= a  
 PN US5837538-A.  
 PD 17-NOV-1998.  
 PE 06-OCT-1995; 540406.  
 PR 06-OCT-1995; US-540406.  
 PR 07-OCT-1994; US-319745.  
 PA (STRD) UNIV LELAND STANFORD JUNIOR.  
 PI Goodrich LV, Johnson RL, Scott MP;  
 DR MPI: 99-023461/02.  
 DR P-PSDB: W72968.  
 PT Nucleic acid encoding vertebrate patched protein and related  
 PT transformants - used to express poly-peptide(s), useful for  
 PT diagnosis and treatment of developmental disorders or cancer, and in  
 PT healing of injured tissue  
 PS Claim 1: Column 43-48; 38pp: English.  
 CC The present sequence represents the mouse patched (ptc) gene. Cells  
 CC containing and expressing the ptc gene are useful: (i) for generating  
 CC production of the protein. These in turn are useful: (i) for generating  
 CC antibodies (Ab); and (ii) to screen for specific-binding ligands  
 CC (potential therapeutic agonists and antagonists). The ptc gene, or its  
 CC fragments, are used to isolate related sequences from other mammals; to  
 CC identify mutations (particularly those associated with genetic diseases  
 CC such as spina bifida and other developmental disorders); to monitor  
 CC expression levels in testis (to determine relationship with sperm  
 CC production) and to isolate 5' non-coding sequences (used to study  
 CC embryonic development and to provide regulated expression of proteins).  
 CC The complete gene can be used in gene therapy, including expression of  
 CC antisense molecules, and to generate transgenic animals for studies of  
 CC embryonic development. Ab are used diagnostically to determine the  
 CC ptc protein on cell surfaces and as competitive inhibitors of signal  
 CC transduction through the ptc ligand. Cells that have been engineered to  
 CC express the ptc protein can be used to promote regrowth and healing of  
 CC damaged tissue (e.g. growth of new teeth) and regulation of the ptc  
 CC protein expression may be useful in cancer treatment (it may control the  
 CC Wnt-1 oncogene).  
 SQ Sequence 5187 BP; 1165 A; 1412 C; 1337 G; 1270 T;

Query Match 8.4%; Score 16; DB 1; Length 5187;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 CAAAACACCAAAATG 30  
|||||  
DB 5069 CAAAACACCAAAATG 5054

RESULT 12  
V64099/c  
ID V64099 standard; DNA: 5187 BP.  
AC V64099;  
DT 26-JAN-1999 (first entry)  
DE Precis coenla patched gene;  
KW Patched gene; ptc; diagnosis; treatment; developmental disorder;  
KW cancer; healing; injured tissue; spina bifida; Wnt-1 oncogene;  
KW sperm production; gene therapy; ss.  
OS Precis coenla.  
PN US837538-A.  
PD 17-NOV-1998.  
PF 06-OCT-1995; 540406.  
PR 06-OCT-1995; US-540406.  
PR 07-OCT-1994; US-319745.  
PA (STRD ) UNIV LEIAND STANFORD JUNIOR.  
PI Goodrich LV, Johnson RL, Scott MP;  
DR WPI: 99-023461/02.  
PT Nucleic acid encoding vertebrate patched protein and related  
PT transforms - used to express poly(peptide(s), useful for  
PT diagnosis and treatment of developmental disorders or cancer, and in  
PT healing of injured tissue  
PS Example: Column 21-26; 38BP; English.  
CC The present invention describes vertebrate and invertebrate patched (ptc)  
CC genes. Cells containing and expressing the ptc gene are used for the  
CC recombinant production of the protein. These in turn are useful: (1) for  
CC generating antibodies (Ab); and (11) to screen for specific-binding  
CC ligands (potential therapeutic agonists and antagonists). The ptc gene,  
CC or its fragments, are used to isolate related sequences from other  
CC mammals; to identify mutations (particularly those associated with  
CC genetic diseases such as spina bifida and other developmental disorders);  
CC to monitor expression levels in testis (to determine relationship with  
CC sperm production) and to isolate 5'-non-coding sequences (used to study  
CC embryonic development and to provide regulated expression of proteins).  
CC The complete gene can be used in gene therapy, including expression of  
CC antisense molecules, and to generate transgenic animals for studies of  
CC embryonic development. Ab are used diagnostically to determine the  
CC ptc protein on cell surfaces and as competitive inhibitors of signal  
CC transduction through the ptc ligand. Cells that have been engineered to  
CC express the ptc protein can be used to promote regrowth and healing of  
CC damaged tissue (e.g. growth of new teeth) and regulation of the ptc  
CC protein expression may be useful in cancer treatment (it may control the  
CC Wnt-1 oncogene). The present sequence represents Precis coenla  
CC (butterfly) patched gene, from the present invention.  
CC Sequence 5187 BP; 1165 A; 1412 C; 1337 G; 1270 T;  
SQ

Query Match 8.4%; Score 16; DB 1; Length 5187;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 CAAAACACCAAAATG 30  
|||||  
DB 5069 CAAAACACCAAAATG 5054

RESULT 13  
V87937  
ID V87937 standard; cDNA: 529 BP.  
AC V87937;  
DT 12-FEB-1999 (first entry)  
DE EST clone Tg708.

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
OS Homo sapiens.  
PN M09845437-A2.  
PD 15-OCT-1998.  
PF 10-APR-1998; U06956.  
PR 10-APR-1997; US-837312.  
PA (GENY ) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racie LA, Spaulding V, Treacy M;  
DR WPI: 99-070078/06.  
PT New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries  
PS Claim 1; Page 229; 641BP; English.  
CC The present sequence represents an expressed sequence tag (EST), and is  
CC a polynucleotide of the invention. The polynucleotides of the invention  
CC are all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The EST sequences are also stated to be useful for gene  
CC therapy  
SQ Sequence 529 BP; 127 A; 137 C; 133 G; 132 T;

Query Match 8.4%; Score 16; DB 1; Length 529;  
Best Local Similarity 100.0%; Pred. No. 9.6;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 TCCAATCCCGGCAT 53  
|||||  
DB 405 TCCAATCCCGGCAT 420

RESULT 14  
X41096  
ID X41096 standard; cDNA: 438 BP.  
AC X41096;  
DT 17-JUN-1999 (first entry)  
DE Human secreted protein 5' EST SEQ ID NO:40.  
KW Human secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN M09906548-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1222.  
PR 01-AUG-1997; US-905135.  
PA (GENST ) GENSET.  
PI Duclert A, Dumas M,Line Edwards J, Lacroix B;  
DR WPI: 99-153778/13.  
PT P-PSDB; Y12263.  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
PT kidney, lung, umbilical cord, placenta and colon tissue  
PS Claim 1; Page 188-189; 824BP; English.  
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also



CC be used to develop products for diagnosis and therapy. The proteins  
 CC obtained may have cytokine activity, cell proliferation/differentiation  
 CC activity, hematopoiesis regulating activity, tissue growth regulating  
 CC activity, reproductive hormone regulating activity, chemotactic/  
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
 CC or other activities. The products can be used in forensic, gene therapy  
 CC and chromosome mapping procedures. The sequences can also be used for  
 CC obtaining corresponding promoter sequences. The nucleic acids encoding  
 CC the signal peptide can be used for directing extracellular secretion of  
 CC a polypeptide or the insertion of a polypeptide into a membrane, or  
 CC importing a polypeptide into a cell.  
 SQ Sequence 438 BP; 106 A; 104 C; 111 G; 115 T;

Query Match 8.4%; Score 16; DB 1; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 9.6;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 TCCAAATCCAGGCAT 53  
 |||||  
 Db 398 TCCAAATCCAGGCAT 413

## RESULT 15

N91208/c  
 ID N91208 standard; DNA; 1998 BP.  
 AC N91208;  
 DT 11-JUN-1990 (first entry)  
 DE Acetyl-CoA reductase gene.  
 KW Polyester biopolymers; acetyl-CoA reductase; polyhydroxybutyrate; ss.  
 OS Zoogloea ramifera strain I-1b-M.  
 FH Key Location/Qualifiers  
 FT cds 37..858  
 FT /\*tag= a  
 PN /product=acetyl-CoA reductase  
 PD MO8900202-A.  
 PF 12-JAN-1989.  
 PF 27-JUN-1988; U02227.  
 PR 29-JUN-1987; US-067695.  
 PA (MASI) Massachusetts Institute of Technology.  
 PI Peoples OP, Sinskey AJ;  
 DR WPI: 89-039655/05.  
 DR P-P8DB; P94155.  
 PT Constructing new polyester biopolymers - using genes encoding  
 PT beta-ketothiolase(s), acetoacetyl-CoA reductase(s) and  
 PT polyhydroxybutyrate synthetase(s).  
 PS Disclosure: fig 6; 67pp; English.  
 CC Gene and the enzyme produced on its expression are used to study the  
 CC polyhydroxybutyrate (PHB) biosynthetic pathway. The enzyme catalyses the  
 CC reduction of acetoacetyl CoA to form D(-)-beta-hydroxybutyryl-CoA, the  
 CC substrate for PHB synthetase. It can be used to control biopolymer  
 CC syntheses and produce new biopolymers.  
 CC See also N91207 and N91209.  
 SQ Sequence 1998 BP; 416 A; 605 C; 637 G; 340 T;

Query Match 7.9%; Score 15; DB 1; Length 1998;  
 Best Local Similarity 100.0%; Pred. No. 36;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 CCAATGCTACCTGC 145  
 |||||  
 Db 55 CCAATGCTACCTGC 41

Search completed: August 5, 1999, 17:40:47  
 Job time: 6282 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:32:41 ; Search time 3075.15 Seconds  
(without alignments)  
153.946 Million cell updates/sec

Title: US-09-049-696-3

Perfect score: 240  
Sequence: 1 AAAAGCTGAGTCTCTGCTT.....GAATTCCTACTATCCAAATG 240

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	92.6	38.6	638	45	AI390317	AI390317 mx03b07.y
2	71	29.6	428	30	Z84017	Z84017 SZ84017.Po
3	32.4	13.5	414	24	H66313	H66313 y72h06.r1
4	32.2	13.4	307	23	H22213	H22213 y138905.r1
5	32.2	13.4	326	25	N49046	N49046 y78d09.r1
6	32.2	13.4	441	29	AA158712	AA158712 z063c05.s
7	32.2	13.4	474	33	AA442061	AA442061 zw63a03.s
8	32.2	13.4	354	49	AI648693	AI648693 tx65g02.x
9	31.2	13.0	446	26	W69657	W69657 z4d8a02.s1
10	31.2	13.0	353	30	AA223445	AA223445 z706c05.r
11	31.2	13.0	406	39	AA828880	AA828880 Od74h02.s
12	31.2	13.0	838	45	AU005029	AU005029 AU005029
13	31.2	12.9	357	34	AA517606	AA517606 v93d05.r
14	31.2	12.9	551	37	AA691227	AA691227 vs14b05.r
15	30.6	12.8	790	45	AU005108	AU005108 AU005108
16	30.6	12.8	897	45	AU006443	AU006443 AU006443
17	30.4	12.7	226	20	Z17523	Z17523 ACTS0097.AC
18	30.4	12.7	338	22	R25739	R25739 yH43D03.r1
19	30.4	12.7	566	48	AI575062	AI575062 UI-R-GO-u
20	30.2	12.6	429	26	W80546	W80546 z482c02.r1
21	30.2	12.6	425	38	AA774566	AA774566 a127c07.s
22	30	12.5	253	23	D60998	D60998 HUM145H1B
23	30	12.5	471	29	AA128853	AA128853 Z113d11.r
24	30	12.5	302	31	AA298602	AA298602 EST114417
25	30	12.5	710	45	AU005203	AU005203 AU005203
26	29.8	12.4	376	36	C67786	C67786 C67786 yxj1
27	29.8	12.4	559	46	AI401905	AI401905 GH03305.3
28	29.6	12.3	360	35	C39391	C39391 C39391 yxj1
29	29.6	12.3	470	39	AA876369	AA876369 OJ24d02.s
30	29.4	12.2	435	31	AA308678	AA308678 EST179555
31	29.4	12.2	411	37	AA711207	AA711207 vt70a10.r
32	29.4	12.2	522	38	AA736449	AA736449 zh31c10.s
33	29.4	12.2	442	44	AI290263	AI290263 qm01c01.x
34	29.4	12.2	419	45	AI360208	AI360208 qy84b09.x
35	29.2	12.2	381	21	T96275	T96275 ye10h05.s1
36	29.2	12.2	430	22	R82389	R82389 y118c06.r1
37	29.2	12.2	460	36	C72859	C72859 C72859 Rice
38	29.2	12.2	491	44	AI267278	AI267278 ag63a12.x
39	29	12.1	289	32	AA348760	AA348760 EST55260
40	29	12.1	410	34	AA492775	AA492775 v176h11.r
41	29	12.1	358	40	AA928536	AA928536 cm73f06.s
42	29	12.1	423	40	AU004108	AU004108 AU004108
43	28.8	12.0	641	41	AI055435	AI055435 coau0003P
44	28.8	12.0	216	50	AV045442	AV045442 AV045442
45	28.8	12.0				

#### ALIGNMENTS

RESULT 1  
AI390317  
LOCUS AI390317 638 bp mRNA  
DEFINITION mx03b07.y1 Scares mouse NML Mus musculus cDNA clone IMAGE:679093 5'  
similar to TR:O18744 O18744 LD-ECAM-1. [3] TR:O18742 TR:O18741 ;  
ACCESSION AI390317  
NID 94216324

VERSION AI390317.1 GI:4216324  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 AUTHORS 1 (bases 1 to 638)  
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritzer, E., Kohn, S., Shan, T., Jackson, T., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
 The Mashu-NCI Mouse EST Project 1999  
 Unpublished (1999)  
 On Aug 26, 1998 this sequence version replaced gi:3478421.

TITLE JOURNAL  
 COMMENT Contact: Marra M/Mashu-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL: contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 This read is a RESEQUENCE of a previously sequenced mouse clone  
 This read has been verified (found to hit its original self in the correct orientation)  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 461.  
 Location/Qualifiers  
 1..638

FEATURES  
 SOURCE  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="11"  
 /clone="IMAGE:679093"  
 /clone\_lib="Soares mouse NML"  
 /tissue\_type="Liver"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCACATCTGAGTGGAGGAGCGGCCGCAATCTTTTCTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 180 a 167 c 145 g 144 t 2 others  
 ORIGIN

Query Match 38.6%; Score 92.6; DB 45; Length 638;  
 Best Local Similarity 61.6%; Pred. No. 1.3e-19;  
 Matches 146; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

OY 1 AAAATGCTGATGTTCTGTTCTGAGTCTACTCTCCAGGTAATGAAACCTTAACG 60  
 DB 343 ACAAGCAGACGTATGATGTCGATGCTCACCCTGCAACATGAGACGACCCCTACACCC 402  
 OY 61 AGCAGATGGCACTGTGGAGAGAGAGGTAAGATCCACCTCATCTCGATTTCATG 120  
 DB 403 TTGATGATGACAGTGGGGGACAGAGACAGTACATACCTTCACTCCAACTTCTTAC 462  
 OY 121 CAGGAAAAAGTTAGCTGATGATGACACCAAGTAGGGGCAATTTGCCATGATGGGCTC 180  
 DB 463 TCACCTGATTAATCTGCGATCTAGGAGACCCGAGGACAGAGCTCTGTCACATGATGGGCC 522  
 OY 181 ATCTAGCAGTGGAGTATTTGACGAGTACATATGATGAGAAATCTTACTTCA 237  
 DB 523 ATCTCGGATGAGATTTGATGAGTATTAAGCTGACGACCCCTTCTCATGATCTCA 579

RESULT 2  
 LOCUS 284017  
 DEFINITION SS284017 Porcine small intestine cDNA library Sus scrofa cDNA clone c12a03 5', mRNA sequence.  
 ACCESSION 284017  
 NID 91806336  
 VERSION 284017.1 GI:1806336  
 KEYWORDS EST.  
 SOURCE Sus scrofa  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 REFERENCE 1 (bases 1 to 428)  
 Wintero, A.K., Fredholm, M. and Davies, W.  
 Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones  
 Mamm. Genome 7 (7), 509-517 (1996)  
 JOURNAL 96327607  
 MEDLINE  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1404772.

FEATURES  
 SOURCE  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone="c12a03"  
 /clone\_lib="Porcine small intestine cDNA library"  
 /note="directionally cloned cDNA in X1-blue MRF"  
 Location/Qualifiers  
 1..428

BASE COUNT 134 a 87 c 100 g 102 t 5 others  
 ORIGIN

Query Match 29.6%; Score 71; DB 30; Length 428;  
 Best Local Similarity 79.0%; Pred. No. 9.9e-13;  
 Matches 83; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 1 AAAATGCTGATGTTCTGTTCTGAGTCTACTCTCCAGGTAATGAAACCTTAACG 60  
 DB 324 AAAATGCTGATGTTCTGTTCTGAGTCTACTCTCCAGGTAATGAAACCTTAACG 383  
 OY 61 AGCAGATGGCACTGTGGAGAGAGGTAAGATCCACCTCA 105  
 DB 384 AGCAGATGGGNAACCTTGTGAGAGAAAGGAGGTAAGATTTATTTC 428

RESULT 3  
 LOCUS H66313  
 DEFINITION yr12h06.r1 Soares fetal liver spleen INF1S Homo sapiens cDNA clone IMAGE:210875 5', mRNA sequence.  
 ACCESSION H66313  
 NID 91025053  
 VERSION H66313.1 GI:1025053  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 414)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Marra, M., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rongling, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 TITLE JOURNAL  
 MEDLINE 97044478



```

VERSION      N49046.1  GI:1190212
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 326)
AUTHORS      Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
             Holtman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
             Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
             Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
             Wilson,R.
TITLE        The Mashu-Merck EST Project
JOURNAL      Unpublished (1995)
COMMENT      On Apr 14, 1993 this sequence version replaced gi:692453.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: T7
High quality sequence stop: 297.
Location/Qualifiers
1..326
/organism="Homo sapiens"
/db_xref="GDB:3898041"
/db_xref="taxon:9606"
/clone IMAGE:27965"
/clone_id="Soares_multiple_sclerosis_2NBHMS"
/sex="male"
/tissue_type="multiple sclerosis lesions"
/dev_stage="Age 46"
/lab_host="DH10B (ampicillin resistant)"
/note="vector: pT73b (Pharmacia) with a modified
polylinker V_type: phagemid; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - Oligo(dt)
primer [5',
TGTTCACATCTGAAGTGGAGCGCGCCGACATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH). "
BASE COUNT   107 a 39 c 66 g 111 t 3 others
ORIGIN
Query Match 13.4%; Score 32.2; DB 25; Length 326;
Best Local Similarity 52.6%; Pred. No.2.4;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

OY 108 CCTGATTCATTGCAGAAAAAGTTAGCTGAATATGACCAACAGTGAGGCATTGTTC 167
DB 100 CTTGGTATTATTAGTAGTGAAGATGAGATGAATGTTTATTATTATTAGGTGAACACATTGC 159
OY 168 CATTGATGGCGCATCTAGCATGGGAGATTTTGTAGCACTACAAATATGATGAGAAATTC 227
DB 160 CCAGGCTGCTTCAATCTCAATTTGGGTTAATGTTATGAAATAAATATGATATAACAC 219
OY 228 TACTATTCCAATG 240
DB 220 TAACATAATAAGG 232

RESULT 6
AA158712/C

```

```

LOCUS       AA158712             441 bp      mRNA           EST            09-MAR-1998
DEFINITION  z63c05.s1 Stratiagene pancreas (#937208) Homo sapiens cDNA clone
IMAGE       IMAGE:591560 3', mRNA sequence.

ACCESSION   AA158712
NID         AA158712
VERSION     AA158712.1
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 441)
AUTHORS    Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
            Kitzman,D., Kuwabara,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,T., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
            WashU-NCI Human EST Project
COMMENT     Unpublished (1997)
            On Sep 12, 1996 this sequence version replaced gi:1397391.

FEATURES             Source
Source
1..441
/organism="Homo sapiens"
/db_xref="GDB:4622946"
/db_xref="taxon:9606"
/clone="IMAGE:591560"
/clone_1db="Stratiagene pancreas (#937208)"
/lab_host="SOLR cells (Kanamycin resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-, Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor
sequence: 5' GAATTCGGCAGCG 3' -3' adaptor sequence: 5'
CTGCAGTGTCTTTTCTTTTTTTT 3'"

BASE COUNT      124 a          102 c          71 g          143 t          1 others
ORIGIN
Query Match      13.4%; Score 32.2; DB 29; Length 441;
Best Local Similarity 52.6%; Prid. No. 2.6;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

CY 108 CCTGATTTCATGCGAGAAAAAAGTAGCGAATATGACACACAAGSTAGGCATTGTGTC 167
Db 159 CTGGATTATTAAGATGATGACGATGCAAAAGTTATTTATTTATTAAGTGGAAGCATTTTGAC 100
CY 168 CATGAGTGGGCTCATCTACGATGGGGAGTATTGACGACAGTACATAATGATGAGAAAATTC 227
Db 99 CCAAGGTGTCTTCATCTTCATATTGGGTTAATGTATAGAAATTAATGATGATTAACATC 40
CY 228 TACTTATCCAAATG 240
Db 39 TAACATAATTAAGG 27

RESULT 7
AA442061/c 474 bp mRNA EST 02-JUN-1997
LOCUS       AA442061
DEFINITION  zw63a03.s1 Soares-total.fetus_NB2HF8_9w Homo sapiens cDNA clone
IMAGE       IMAGE:774700 3' similar to SW:OBP2.DICD1 P36413 DIHYDROLIPOAMIDE
ACETYLTRANSFERASE COMPONENT 1, mRNA sequence.
ACCESSION   AA442061

```

NID 92153939  
VERSION AA442061.1 GI:2153939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 474)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R., and Wilson, R.  
TITLE WashU-Merck EST Project 1997  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397911.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -41m3 fwd. ET from AmerSham  
High quality sequence stop: 429.  
Location/Qualifiers  
1. 474  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:774700"  
/clone\_1lb="Soares\_total\_fetus\_MB2HF8\_9w"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I. Site\_2: Eco RI. 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - Oligo(dt) primer [5' TGTTCACATCTGAGTGGAGCGGCGGCTTAATTTTATTTTATTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 134 a 111 c 78 g 151 t

Query Match 13.4%; Score 32.2; DB 33; Length 474;  
Best Local Similarity 52.6%; Pred. No. 2.7;  
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 108 CCTGATTTCATTCAGAGAAAAAGTTAGCTGAATGACACACAGGTAGGCAATTTGTC 167  
Db 146 CTTGATTTTAACTATGAGAGTGAATGTTTATTTTAAAGTGAAGCAATTTGAC 87  
QY 168 CATGAGTGGGCTCATCTACAGTGGAGTATTGACAGTACATATGATGAGAAATTC 227  
Db 86 CCAAGGTGCTTCTCATCTTCAATTTGGGTTAATGTTTAAATAATGATGATAAATC 27  
QY 228 TACTATCCATG 240  
Db 26 TAACATAAAG 14

RESULT 8  
LOCUS AI648693 354 bp mRNA EST 30-APR-1999  
DEFINITION tx65g02.x1 NCI-CGAP\_Ut1 Homo sapiens cDNA clone IMAGE:2274482 3' similar to SM:ODPX\_HUMAN 00030 PYRUVATE DEHYDROGENASE PROTEIN X COMPONENT PRECURSOR ;, mRNA sequence.  
ACCESSION AI648693

NID 94729527  
VERSION AI648693.1 GI:4729527  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 354)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Feb 22, 1999 this sequence version replaced gi:4282908.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.llnl.gov/dbp/image/image.html  
Seq primer: -40UP from gibco  
High quality sequence stop: 351.  
Location/Qualifiers  
1. 354  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2274482"  
/clone\_1lb="NCI-CGAP\_Ut1"  
/tissue\_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6. Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 117 a 52 c 74 g 111 t

Query Match 13.4%; Score 32.2; DB 49; Length 354;  
Best Local Similarity 52.6%; Pred. No. 2.5;  
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 108 CCTGATTTCATTCAGAGAAAAAGTTAGCTGAATGACACACAGGTAGGCAATTTGTC 167  
Db 186 CTTGATTTTAACTATGAGAGTGAATGTTTATTTTAAAGTGAAGCAATTTGAC 245  
QY 168 CATGAGTGGGCTCATCTACAGTGGAGTATTGACAGTACATATGATGAGAAATTC 227  
Db 246 CCAAGGTGCTTCTCATCTTCAATTTGGGTTAATGTTTAAATAATGATGATAAATC 305  
QY 228 TACTATCCATG 240  
Db 306 TAACATAAAG 318

RESULT 9  
LOCUS W69657 446 bp mRNA EST 16-OCT-1996  
DEFINITION zd48a02.s1 Soares fetal heart\_NbHH19W Homo sapiens cDNA clone IMAGE:343850 3', mRNA sequence.  
ACCESSION W69657  
NID 91378987  
DEFINITION W69657.1 GI:1378987  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

RESULT	10		
AA223445			
LOCUS			
DEFINITION	AA223445	353 bp	mRNA
	str6c05.r1		EST
	CDNA clone IMAGE:650696	5'	12-MAR-1998
	PHOSPHATASE 2C (HUMAN).,		Homo sapiens
			similar to gp.D13540 PROTEIN-TYROSINASE
ACCESSION	AA223445		
NID	91843970		
VERSION	AA223445.1	GI:1843970	
KEYWORDS	EST.		
SOURCE	human.		

RESULT	11						
LOCUS	AA828880						
DEFINITION	AA828880	406 bp	mRNA	EST	07-APR-1998		
	o374h02.s1	NCI_CGAP_Ov2	Homo sapiens	CDNA clone	IMAGE:1373715,	mRNA	
ACCESSION	AA828880						
NTID	92901979						
VERSION	AA828880.1	GI:2901979					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;						
	Eutheria; Primates; Catarrhini; Homiidae; Homo.						
REFERENCE	1 (bases 1 to 406)						
AUTHORS	NCI-CGAP	<a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .					
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),						
	Tumor Gene Index						



JOURNAL  
COMMENT Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2045865.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bdrip/image/image.html

Insert Length: 1249 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 396.

## FEATURES

Location/Qualifiers

1..406  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="3q"  
/clone="IMAGE:1373715"  
/clone\_lib="NCI-CGAP\_Ov2"  
/sex="female"  
/tissue\_type="ovary"  
/lab\_host="DH10B"  
/note="Vector: PAMPI0; mRNA made from invasive ovarian  
tumor, CDNA made by oligo-dt priming. Non-directionally  
cloned. Size-selected on agarose gel, average insert size  
600 bp. Reference: Kitzman et al. (1996) Cancer Research  
56:5380-5383."

BASE COUNT 121 a 85 c 79 g 121 t  
ORIGIN

Query Match 13.0%; Score 31.2; DB 39; Length 406;  
Best Local Similarity 55.6%; Pred. No. 5.4;  
Matches 60; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

OY 63 CAGATGGCACTGTGGAGGAGGTGAGGACCTGACCTGATTTCATTGCA 122  
DB 39 CAAATAGAGAAATAGCGTTAGAACGCTCAATCATTTAATTAACCCAGATCTCTTACC 98  
OY 123 GGAAGAAAGTTAGCTGAATATGACCAACAAGTAGGCGATTGTCCAT 170  
DB 99 CTTAGTTTACGATGAATTTGAACACAATTTGCCCAATTGTCCCT 146

RESULT 12  
AU005029 838 bp mRNA EST 19-JAN-1999  
LOCUS AU005029 Bombyx mori p50(Daizo) Bombyx mori cDNA clone vs30160,  
DEFINITION mRNA sequence.  
ACCESSION AU005029  
NID 94162400  
VERSION AU005029.1 GI:4162400  
KEYWORDS EST.  
SOURCE domestic silkworm.  
ORGANISM Bombyx mori

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
Ditrysia; Bombycoidea; Bombycidae; Bombyx.

REFERENCE 1 (bases 1 to 838)  
Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.  
Establishment of cDNA database of Bombyx mori

JOURNAL Unpublished (1999)  
COMMENT On Oct 20, 1997 this sequence version replaced gi:2520670.

CONTACT: Mita K  
Genome Research Group  
National Institute of Radiological Sciences

Anagawa 4-9-1, Inage, Chiba 263-8555, Japan  
Email: kmitsukawa@nirs.go.jp  
PROJECT = 'CREST project by JST'.  
Location/Qualifiers

## FEATURES

1..838  
/organism="Bombyx mori"  
/strain="p50(Daizo)"  
/db\_xref="taxon:7091"  
/clone\_lib="Bombyx mori p50(Daizo)"  
/clone="ws30160"

BASE COUNT 233 a 185 c 195 g 222 t 3 others  
ORIGIN

Query Match 13.0%; Score 31.2; DB 45; Length 838;  
Best Local Similarity 54.5%; Pred. No. 6.6;  
Matches 60; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 38 AGCTATGATGAAACCTACACGACGATGGCAACTGTGAGAGAGGATGAAAGCAT 97  
DB 115 AGCTGAGATGAGCCGACGCTGATGTCGACAGGTAGACCCCGCATGAATATT 174  
OY 98 CCACCTCACTCCTGATTTCATTCAGAGAAAGTTAGCTGAATATGAC 147  
DB 175 CCGTCTCACACCAAGNANCTGCCGATGTCGTCAGGTGACATAGTC 224

RESULT 13  
AA517606/c 357 bp mRNA EST 14-JUL-1997  
LOCUS AA517606/c  
DEFINITION v93d05.r1 Knowles Solter mouse embryonic stem cell Mus musculus  
CDNA clone IMAGE:894537.5, similar to gb:M44739 CALRETICULIN  
PRECURSOR (HUMAN); gb:X14926 Mouse mRNA for calreticulin (MOUSE);  
mRNA sequence.

ACCESSION AA517606  
NID 92257130  
VERSION AA517606.1 GI:2257130  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 357)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucada, T., Lacy, M., Le, M., Martin, U., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:693326.

Contact: Marras M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:522497  
Trace considered overall poor quality  
High quality sequence stop: 1.

## FEATURES

Location/Qualifiers

1..357  
/organism="Mus musculus"  
/strain="B6D2 F1/J"  
/db\_xref="taxon:10090"  
/map="11"  
/clone="IMAGE:894537"  
/clone\_lib="Knowles Solter mouse embryonic stem cell"  
/dev\_stage="embryo"

```

BASE COUNT      97 a      81 c      97 g      82 t
ORIGIN
/lab host="DH10B"
/notes="vector: pSPORT; site.1: NotI; site.2: SalI; cloned
unidirectionally from mRNA prepared from 800 blastocysts.
Primer: SalI(dt): 5'-CGGTCAACGTCACCGTATTTTITTTT-3'
cDNAs were cloned into the NotI/SalI sites of a pSPORT
vector (Life Technologies)."

```

Query Match	12.98;	Score 31;	DB 34;	Length 357;
-------------	--------	-----------	--------	-------------

Matches 43; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY	58	CTGAGCAGATGGCCAACTGTGAGAGAAGGGTGAAGACTCACCTACCTCTCATTTCA	117
Dd	216	CTGACCAGATGGGTCAAGTGTATTAGGTGTGTGAACATCATTCCTTAACACCGGATATCA	157
QY	118	TTC	120
Dd	156	TTG	154

RESULT	14
AA691227	
LOCUS	551 bp mRNA EST 16-DEC-1997
DEFINITION	vsl4b05.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus CDNA clone IMAGE:118161 5', mRNA sequence.

ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM
AA691227	92692163	AA691227.1	GI:2692163	EST.	house mouse.
					Mus musculus

REFERENCE	1 (bases 1 to 551)
AUTHORS	Marrà,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1397428

Contact: Marra M/Mouse EST Project

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Fax: 314 286 1810

This clone is available royalty-

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

vector to vector length is 577

High quality sequence stop: 513.

FEATURES	Location/Qualifiers
SOURCE	1. .551

```
/strain="FVB/N"
```

```

/mmap="17q21"

```

```
/clone_lib="Barstead m
```

```
/Lab_host="DH10B"
```

polylinker; site\_1:

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - oligo(dT) primer  
 5'GTGATGACATCTGTGAAGGGAGAGCGCCGCTTTTTTTTTTTTTTTTTTT  
 T3'] double-stranded cDNA was ligated to Eco RI  
 adaptors [AATTCGATCCTGG], digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT73  
 vector. Library constructed by Bob Barstead. "

Query Match	12.98; Score 31; DB 37; Length 551;
-------------	-------------------------------------

Matches	73;	Conservative	0;	Mismatches	70;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

QY	38	AGGTAATGATGAACCCCTACACTGACGACGATGGGCAATGTGGAGAAAGGTTGAAAGGAT	97
Db	177	AGTTAATATACCCATGCCCAAAATATGAAATCTTGTACTGTGTGGGGGAGAGGACGATTTGAC	236
QY	98	CCACCTCACTCCTGATTCATTGTCAGGAAAAAGTTAGTCGAATATGACACCAAGGTAG	157
Db	237	CCCTTTCAGGCACTTTTGCATGTGTTTTGATCAAGTCCCAAGGTATATGTGACAGGTTG	296

Qy 158 GGCATTGTCCATGAGTGGGCTC 180  
 | | | | | | | |  
 Db 297 TGGTCTGATCCGTGAGTAGGTTT 319

RESULT	15
LOCUS	AF0005108
DEFINITION	AF0005108 Bombyx mori 790 bp mRNA EST 19-JAN-1999
	AF0005108 Bombyx mori p50(DaIzo) Bombyx mori cDNA clone w30259
	mRNA sequence.

```

VERSION      AU005108.1  GI:4162479
KEYWORDS
SOURCE       EST.
              domestic silkworm.

```

REFERENCE  
1. (bases 1 to 790)  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta  
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
Ditrysia; Bombycoidea; Bombycidae; Bombyx.  
1. (bases 1 to 790)

AUTHORS Mita, K., Moriyasu, M., Shinada, T., Okano, K. and Maeda, S  
 TITLE Establishment of cDNA database of Bombyx mori  
 JOURNAL Unpublished (1999)  
 COMMENT On Jan 14, 1998 this sequence version replaced gi:179

Contact: Mlta K

National Institute of Radiological Sciences  
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

PROJECT - 'CREST project by JST'.

FEATURES  
source

```

/strain="p50(Daizo)"

```

```
/clone="ws30259"
```

BASE COUNT	224 a	172 c	180 g	210 t	4 others
	/clone_11b-"Bombyx mori p50(Daizo)"				

Query Match	12.88; Score 30.6; DB 45; Length 790;
-------------	---------------------------------------

Matches	60;	Conservative	0;	Mismatches	50;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

Oy      38 AGGTATGATGAACCCCTACACTGAGCAGATGGSCAACTTGTGGAGAAGAGGTGAALAGAT    97  
       ||| |||| | | | | | | | | | | | |  
Db      90 AGCTGAGATGAGCGCAACGTCGATCAGTTGTGCAGACGGTAGACC GCCGATGAATAATT    145

Mon Aug 9 13:26:16 1999

us-09-049-696-3.rst

Page 9

QY 98 CCACCTCACTCTGATTTCATTCAGGAAAAAGTTAGTGAATATGAC 147  
|| ||||| || ||||| || ||||| || |||||  
Db 150 CCCTTCACACCCGAAGAGACTGCCGATGTCGTCAAGTCACTAGGTC 199

Search completed: August 6, 1999, 13:32:46  
Job time: 12964 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:39:27 ; Search time 650.92 Seconds  
(without alignments)  
92.248 Million cell updates/sec

Title: US-09-049-696-3  
Perfect score: 240  
Sequence: 1 AAAATGCTGATGTTCTGTT.....GAAATTCATTATCCAAAG 240

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	7.9	710	1 006308	Sequence of DNA fr
2	19	7.9	2184	1 006301	Sequence encoding
3	18	7.9	447	1 V90187	EST Clone DH318. N
4	18	7.5	735	1 Q73229	Soluble human inte
5	17	7.1	1560	1 Q29613	Encodes carp FLA1a
6	17	7.1	116624	1 V52850	Human eyal gene co
7	17	7.1	1686	1 X52240	Protein PRO229 CDN
8	15	6.2	9320	1 N91779	Sequence encoding
9	15	6.2	1990	1 N92451	CDNA insert of vec
10	15	6.2	2069	1 N92452	CDNA insert of vec
11	15	6.2	2400	1 Q49353	deg-3. Long distan
12	15	6.2	30	1 086486	Primer, e22, for a
13	15	6.2	110000	1 T58840_4	Continuation (5 of
14	15	6.2	27	1 T48019	Encodes FIV princ
15	15	6.2	7918	1 T60450	Melanoma-associate
16	15	6.2	680	1 T43268	Human fibroblast g
17	15	6.2	1979	1 T87015	Human Fc-gamma rec
18	15	6.2	2125	1 T87016	Human Fc-gamma rec
19	15	6.2	961	1 T86313	Human fibroblast g
20	15	6.2	2700	1 V21641	Anabaena sucrose s
21	15	6.2	961	1 V2187	Fibroblast growth
22	15	6.2	625	1 V40885	Coding sequence of
23	15	6.2	12566	1 V52282	Streptococcus pneu
24	15	6.2	13121	1 V52359	Streptococcus pneu
25	15	6.2	1755	1 V43025	Streptococcus pneu
26	15	6.2	814	1 T98772	DNA encoding a S.
27	15	6.2	1692	1 V53138	Helicobacter lysyl
28	15	6.2	2007	1 V71089	dnax gene of Therm
29	15	6.2	7932	1 X13194	Enterococcus faeca
30	15	6.2	9121	1 X13158	Enterococcus faeca
31	15	6.2	3852	1 X13071	Enterococcus faeca
32	15	6.2	32768	1 X13060	Enterococcus faeca
33	15	6.2	126	1 X12177	Human biallelic po
34	15	6.2	645	1 X04348	Human secreted pro
35	15	6.2	491	1 X20972	Polynucleotide seg
36	15	6.2	53585	1 X20251	Borrelia burgdorfe
37	15	6.2	3739	1 X18959	Non-B, non-C, non-
38	15	6.2	631	1 X18988	Non-B, non-C, non-
39	15	6.2	4516	1 X26856	A Brissica napus g
40	15	6.2	3441	1 X40160	Gastric cancer ass
41	15	6.2	1929	1 X08426	Acidic leucine aml
42	14	5.8	941	1 002893	CDNA insert of vec
43	14	5.8	4434	1 N91325	DNA encoding human

44 14 5.8 735 1 N91020 Nicotiana tabacum  
45 14 5.8 250 1 X12585 Human biallelic po

ALIGNMENTS

```
RESULT 1
ID 006308/c 006308 standard; DNA; 710 BP.
AC 006308;
DT 29-JAN-1991 (first entry)
DE Sequence of DNA fragment F7 of the human IFN-gamma receptor.
KW IFN-gamma receptor; autoimmune disease; multiple sclerosis;
  hypersensitivity; ds.
OS Homo sapiens.
PN EP-393502-A.
PD 24-OCT-1990.
PF 11-APR-1989; EP-810295.
PR 19-APR-1989; EP-810295.
PA Fountoulakis M, Garotta G, Stuber D;
  (HOFF) HOFFMANN-LA ROCHE AG.
DR WPI: 90-322042/43.
P-PSDB: R07472.
PT Soluble interferon-gamma receptors - for treating auto-immune
  diseases; chronic inflammations, etc.
PS Disclosure; Fig 18; 174pp; English.
CC IFN-gamma is a therapeutically active agent in the treatment
  of autoimmune disease, allograft transplant rejections, multiple
  sclerosis, chronic inflammations and delayed hypersensitivity. It is
  also useful in identifying IFN-gamma agonists and antagonists.
  See also 006301.
SQ Sequence 710 BP; 206 A; 147 C; 166 G; 191 T;

Query Match 7.9%; Score 19; DB 1; Length 710;
Best Local Similarity 100.0%; Pred No. 0.77;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 TACAAATGATGATGGAAT 225
  |||||
Db 322 TACAAATGATGGAAT 304

RESULT 2
ID 006301/c 006301 standard; DNA; 2184 BP.
AC 006301;
DT 28-JAN-1991 (first entry)
DE Sequence encoding SacI/Asp718I fragment of plasmid pBABLU carrying
  human interferon-gamma receptor gene.
KW IFN-gamma receptor; autoimmune disease; multiple sclerosis;
  hypersensitivity; ds.
OS Homo sapiens.
FH key Location/Qualifiers
FT cds 85..1531
FT EP-393502-A.
FT 24-OCT-1990.
FT 11-APR-1990; 106992.
PR 19-APR-1989; EP-810295.
PA (HOFF) HOFFMANN-LA ROCHE AG.
PI Fountoulakis M, Garotta G, Stuber D;
  WPI: 90-322042/43.
DR P-PSDB: R07469.
PT Soluble interferon-gamma receptors - for treating auto-immune
  diseases; chronic inflammations, etc.
PS Disclosure; Fig 1; 174pp; English.
CC Sequence may be used to transform prokaryotic or mammalian host
  cells via an expression vector, allowing production of the IFN-gamma
  receptor in pure form.
CC The gene product is a therapeutically active agent in the treatment
  of autoimmune disease, allograft transplant rejections, multiple
```

CC sclerosis, chronic inflammations and delayed hypersensitivity. It is  
CC also useful in identifying IFN-gamma agonists and antagonists.  
SQ Sequence 2184 BP; 688 A; 413 C; 451 G; 632 T;

Query Match 7.9%; Score 19; DB 1; Length 2184;  
Best Local Similarity 100.0%; Pred. No. 0.75;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 TACAAATATGATGAGAAAT 225  
DB 340 TACAAATATGATGAGAAAT 322

RESULT 3  
V90187  
ID V90187 standard; CDNA: 447 BP.  
AC V90187;  
DE 15-FEB-1999 (first entry)  
DE EST clone DH318.  
KW Human: secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
KW gene therapy; ss.  
OS Homo sapiens.  
PN MO9845436-A2.  
PD 15-OCT-1998.  
PR 10-APR-1998; U06955.  
PR 10-APR-1998; US-838821.  
PA (GEMT) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racie LA, Spaulding V, Treacy M;  
DR MPI: 99-070077/06.  
PT New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon CDNA libraries.  
PS Claim 1: Page 459; 618pp; English.  
CC The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
SQ Sequence 447 BP; 120 A; 107 C; 95 G; 125 T;

Query Match 7.9%; Score 19; DB 1; Length 447;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 TACAAATATGATGAGAAAT 225  
DB 168 TACAAATATGATGAGAAAT 186

RESULT 4  
O73229/C  
ID O73229 standard; CDNA: 735 BP.  
AC O73229;  
DE 11-APR-1995 (first entry)  
DE Soluble human interferon gamma receptor coding sequence.  
KW Interferon; gamma; IFN; receptor; immunoglobulin; constant domain;  
KW light chain; heavy chain; 19; chimeric protein; fusion protein;  
KW autoimmune disease; chronic inflammation; allotransplant; rejection;  
KW multiple sclerosis; fulminant hepatitis; neurological disease; AIDS;  
KW poliovirus; Lyme disease; septicemia; treatment; therapy;  
KW delayed type hypersensitivity; ss.

OS Homo sapiens.

FH Key Location/Qualifiers  
FT cds 1..735  
FT /tag= a  
FT /product= Soluble Interferon gamma receptor.

FT signal\_peptide 1..51  
FT /tag= b  
FT mat\_peptide 52..735  
FT /tag= c

PN EP-614981-A.  
PD 14-SEP-1994.  
PR 18-FEB-1994; 102452  
PR 05-MAR-1993; EP-810170.  
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.  
PI Dembic Z, Garotta G, Gentz R;  
DR MPI: 94-281208/35.  
DR P-PSDB: R62023.

PT Chimeric human interferon-gamma receptor/immunoglobulin proteins  
PT - used to inhibit binding of interferon-gamma to its specific  
PT receptor in the treatment of illnesses  
PS Disclosure: Figure 1; 29pp; English.  
CC The soluble form of the interferon (IFN) gamma receptor comprises  
CC the whole extracellular domain of the natural receptor from the N-  
CC terminus to the transmembrane region, lacks the cytoplasmic and  
CC transmembrane domains of the natural receptor and specifically binds  
CC IFN-gamma. The sequence encoding the soluble IFN-gamma receptor can  
CC be used in constructs encoding chimeric proteins where the other  
CC component of the chimeric protein is part or all of the constant  
CC domain of a human immunoglobulin heavy or light chain. The  
CC recombinant proteins can be used to inhibit IFN-gamma binding to its  
CC specific receptor. They can be used for the treatment of  
CC illnesses, especially autoimmune diseases, chronic inflammation,  
CC delayed type hypersensitivity, allotransplant rejections, multiple  
CC sclerosis, fulminant hepatitis, inflammatory neurological diseases  
CC and neurological complications of AIDS, poliovirus infections, Lyme  
CC disease and septicemia. The presence of the immunoglobulin  
CC component in the chimeric protein increases the proteins half life in  
CC vivo.

Query Match 7.5%; Score 18; DB 1; Length 735;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 TACAAATATGATGAGAAA 224  
DB 256 TACAAATATGATGAGAAA 239

RESULT 5  
O29613/C  
ID O29613 standard; DNA: 1560 BP.  
AC O29613;  
DE 12-NOV-1992 (first entry)  
DE Encodes carp TIRAPalpha-I  
KW Polymerase chain reaction; detection; intercellular recognition;  
KW adhesion; antigen; migratory fishes; carp; HLA; ss.  
OS Cyprinus carpio  
FH Key Location/Qualifiers  
FT exon 102..336  
FT /tag= a  
FT /label= Exon 1  
FT exon 519..771  
FT /tag= b  
FT /label= Exon 2  
FT exon 1214..1490  
FT /tag= c  
FT /label= Exon 3  
PN J04112790-A.  
PD 04-APR-1992; 230534.  
PR 09-MAR-1990; JP-230534.  
PR 09-MAR-1990; JP-230534.

PA (GAKK-) GAKKO HOJIN FUJITA GAKUEN;  
 PA (MITR ) MITSUI PHARM INC.;  
 PA (MITR ) MITSUI TOATSU CHEM.  
 DR WPI: 92-173146/21.  
 DR P-PSDB: R23862.  
 PT Oligo-nucleotide coding aminoacid sequence of antigen -for  
 inter-cellular recognition and adhesion, for determining river  
 and country of origin of migratory fish  
 PS Disclosure: Fig 3b; 14pp: Japanese.  
 CC This sequence represents the Tlra1alpha-1 gene, which codes  
 the H chain from carp MHC class I molecules. The gene can be  
 probed for using Q24545. The probe can be used in a method of  
 identification which can specify the river and the country where  
 migratory fish were born.  
 CC See also Q24541,3-6, R27686-7, Q29612-3, R23862,4 .  
 CC Sequence 1560 BP; 492 A; 274 C; 316 G; 478 T;  
 SQ

Query Match 7.1%; Score 17; DB 1; Length 1560;  
 Best Local Similarity 100.0%; Pred. No. 8.1;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 208 ACAATATGATGAGAAA 224  
 DB 505 ACAATATGATGAGAAA 489

RESULT 6  
 V52850  
 ID V52850 standard; DNA; 116624 BP.  
 AC V52850;  
 DT 06-NOV-1998 (first entry)  
 DE Human eyal gene contig 4405-9480.  
 KW Human; eyal; branchio-oto-renal syndrome; BOR; carcinogenesis; cancer;  
 KW kidney tumorigenesis; tumour; diagnosis; ss.  
 OS Homo sapiens.  
 PN WO9832849-A2.  
 PD 30-JUL-1998.  
 PF 28-JAN-1998: E00433.  
 PR 29-JAN-1997: US-036579.  
 PA (INSP ) INST PASTEUR.  
 PI Abdelhak S, Compain S, Petit C, Vasiliki K, Vincent C,  
 PI Weil D;  
 DR WPI: 98-427945/36.  
 PT Nucleic acid corresponding to human genes implicated in  
 branchio-oto-renal syndrome - useful for, e.g. diagnosis and  
 treatment of syndrome and possibly some carcinogenic processes,  
 PT particularly in kidneys  
 PS Claim 1, Page 44-105; 191pp: English.  
 CC The present sequence represents a portion of the human eya-1 gene.  
 CC Alterations in the eyal gene are associated with branchio-oto-renal  
 CC (BOR) syndrome, including the form (BO) without renal anomalies, and  
 CC possibly in some carcinogenic processes, particularly in the kidney.  
 CC The related genes eya-2 and eya-3 may also be involved in development  
 CC of tumours and cell differentiation. Antibodies specific for EYA  
 CC proteins are used to detect the proteins by immunassay, while genetic  
 CC alterations linked to BOR are detected by amplifying DNA with primers  
 CC (see V52857 to V52893) then identifying any mutations or deletions. EYA1  
 CC and EYA1-B proteins, corresponding nucleic acid, antisense sequences and  
 CC Ab can be used therapeutically to modulate expression of EYA1(B), or  
 CC their active derivatives, especially for treating renal disease  
 CC associated with eyal abnormalities, also for ensuring correct  
 CC development of grafted organs, especially where these are embryonic.  
 CC Therapeutic proteins are administered to maintain 0.1-10 mu g /ml in  
 CC body fluids, locally or systemically. Nucleic acid is administered in  
 CC usual gene therapy vectors (optionally to cells ex vivo) at doses of  
 CC 0.1-100 mu g.  
 CC Sequence 116624 BP; 35549 A; 20868 C; 23043 G; 37152 T;  
 SQ

Query Match 7.1%; Score 17; DB 1; Length 116624;  
 Best Local Similarity 100.0%; Pred. No. 7.4;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 209 CAATATGATGAGAAAT 225  
 DB 93605 CAATATGATGAGAAAT 93621

RESULT 7  
 X52240/c  
 ID X52240 standard; DNA; 1666 BP.  
 AC X52240;  
 DT 25-JUN-1999 (first entry)  
 DE Protein PRO229 cDNA clone DNA33100-1159.  
 KW Secreted protein; transmembrane protein; human; enterocolitis;  
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;  
 KW congenital microvillus atrophy; skin disease; cell growth;  
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophie areata;  
 KW anti-thrombotic; wound healing; tissue repair; ss.  
 OS Homo sapiens.  
 PN WO9914328-A2.  
 PD 25-MAR-1999.  
 PF 16-SEP-1998: U19330.  
 PR 25-NOV-1997: US-066840.  
 PR 17-SEP-1997: US-059113.  
 PR 17-SEP-1997: US-059115.  
 PR 17-SEP-1997: US-059117.  
 PR 17-SEP-1997: US-059119.  
 PR 17-SEP-1997: US-059121.  
 PR 17-SEP-1997: US-059122.  
 PR 17-SEP-1997: US-059184.  
 PR 18-SEP-1997: US-059263.  
 PR 18-SEP-1997: US-059266.  
 PR 15-OCT-1997: US-062125.  
 PR 17-OCT-1997: US-062285.  
 PR 17-OCT-1997: US-062287.  
 PR 21-OCT-1997: US-063486.  
 PR 24-OCT-1997: US-062814.  
 PR 24-OCT-1997: US-062816.  
 PR 24-OCT-1997: US-063045.  
 PR 24-OCT-1997: US-063120.  
 PR 24-OCT-1997: US-063121.  
 PR 24-OCT-1997: US-063127.  
 PR 24-OCT-1997: US-063128.  
 PR 27-OCT-1997: US-063329.  
 PR 27-OCT-1997: US-063327.  
 PR 28-OCT-1997: US-063541.  
 PR 28-OCT-1997: US-063542.  
 PR 28-OCT-1997: US-063544.  
 PR 28-OCT-1997: US-063549.  
 PR 28-OCT-1997: US-063550.  
 PR 28-OCT-1997: US-063564.  
 PR 29-OCT-1997: US-063435.  
 PR 29-OCT-1997: US-063704.  
 PR 29-OCT-1997: US-063732.  
 PR 29-OCT-1997: US-063738.  
 PR 29-OCT-1997: US-063734.  
 PR 29-OCT-1997: US-064215.  
 PR 29-OCT-1997: US-063735.  
 PR 31-OCT-1997: US-063870.  
 PR 31-OCT-1997: US-064103.  
 PR 03-NOV-1997: US-064248.  
 PR 07-NOV-1997: US-064809.  
 PR 12-NOV-1997: US-065186.  
 PR 17-NOV-1997: US-065846.  
 PR 18-NOV-1997: US-065693.  
 PR 21-NOV-1997: US-066120.  
 PR 21-NOV-1997: US-066364.  
 PR 24-NOV-1997: US-066772.  
 PR 24-NOV-1997: US-066466.  
 PR 24-NOV-1997: US-066770.  
 PR 24-NOV-1997: US-066511.  
 PR 24-NOV-1997: US-066453.

PA (GETH) GENENTECH INC.  
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;  
 DR WPI: 99-229533/19.  
 DR P-PSDB: Y13369.  
 PT New isolated human genes and polypeptides used in, e.g. treatment of  
 PT gastrointestinal ulceration  
 PS Claim 2: Fig 53: 320pp; English.  
 CC X52213-74 encode secreted and transmembrane human proteins, and are  
 CC obtained from cDNA libraries, prepared from fetal lung, fetal kidney,  
 CC fetal brain, fetal liver and fetal retina. The encoded polypeptides  
 CC have specific uses based on their homology to known polypeptides,  
 CC e.g. PRO211 and PRO217 can be used for disorders associated with the  
 CC preservation and maintenance of gastrointestinal mucosa and the repair  
 CC of acute and chronic mucosal lesions (e.g. enterocolitis,  
 CC Zollinger-Ellison syndrome, gastrointestinal ulceration and congenital  
 CC microvillus atrophy), skin diseases associated with abnormal keratinocyte  
 CC differentiation (e.g. psoriasis, epithelial cancers such as lung squamous  
 CC cell carcinoma of the vulva and gliomas), potent effects on cell growth  
 CC and development, diseases related to growth or survival of nerve cells  
 CC including Parkinson's disease, Alzheimer's disease, ALS, neuropathies or  
 CC cancer. PRO265 can be used as for fibromodulin, e.g. for reducing dermal  
 CC scarring. PRO264 can be used as a target for anti-tumor drugs. PRO533  
 CC may be used in the treatment of Usher Syndrome or Atrophia areata;  
 CC PRO269 can be used as an anti-thrombotic agent; PRO287 polypeptides  
 CC and portions may have therapeutic applications in wound healing and  
 CC tissue repair; PRO317 can be used for treating problems of the kidney,  
 CC uterus, endometrium, blood vessels, or related tissue, e.g. in the  
 CC heart of genital tract.  
 SO Sequence 1686 BP; 431 A; 396 C; 451 G; 408 T;

Query Match 7.1%; Score 17; DB 1; Length 1686;  
 Best Local Similarity 100.0%; Pred. No. 8.1;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 CCTACACTGACGAGATG 68  
 DB 1142 CCTACACTGACGAGATG 1126

## RESULT 8

ID N91779 standard; cDNA: 9520 BP.  
 AC N91779;  
 DT 27-JAN-1991 (first entry)  
 DE Sequence encoding mouse laminin A chain  
 KW Peripheral nerve regeneration; ss.  
 OS Mouse.  
 FH Key Location/Qualifiers  
 FT cds 49..9303  
 FT mat\_peptide /\*tag= a  
 FT 121..9303  
 PN US7267564-A.  
 PD 28-FEB-1989  
 PF 07-NOV-1988; 267564.  
 PR 07-NOV-1988; US-267564.  
 PA (USSH) US Dept Health & Human.  
 PI Yamada Y, Sasaki M, Kleinman HK, Martin GR;  
 DR WPI: 89-138175/18.  
 DR P-PSDB: P94758.  
 PT DNA encoding human laminin A chain, used in vector system -  
 PT and new synthetic peptide(s) with laminin-type biological  
 PT activity  
 PS Disclosure; Figure 4; 90pp; English.  
 CC The sequence encoding mouse laminin A chain (N91779) is used as a probe  
 CC to screen a human cDNA library. Laminin is a very potent and rapid  
 CC stimulator of neurite outgrowth and promotes both central and peripheral  
 CC nerve regeneration.  
 SO Sequence 9520 BP; 2376 A; 2924 C; 2167 G; 2035 T; 18 Others;

Query Match 6.2%; Score 15; DB 1; Length 9520;

Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 CTGTGAGAGAGAGG 88  
 DB 189 CTGTGAGAGAGAGG 203

## RESULT 9

ID N92451/c  
 AC N92451;  
 DT 04-MAY-1990 (first entry)  
 DE cDNA insert of vector pCD(SR alpha)-GP5.  
 KW Vector pCD(SR alpha)-GP5; Fc gamma RIII; ss;  
 KW Immune thrombocytopenic purpura.  
 FH Key Location/Qualifiers  
 FT cds 18..719  
 FT /\*tag= a  
 PN EP-343950-A.  
 PD 29-NOV-1989.  
 PF 24-MAY-1989; 305241.  
 PR 27-MAY-1988; US-199513.  
 PR 19-JAN-1989; US-300039.  
 PA (SCHE) Schering Biotech Corp.  
 PI Peltz GA, Moore KM;  
 DR WPI: 89-349994/48.  
 DR P-PSDB: P91540.  
 PT Soluble and membrane-bound Fc-gamma-RIII - useful for treating immune  
 PT thrombocytopenic purpura, and in assays for immune complexes.  
 PS Disclosure; Fig. 1A-B: 25pp; English.  
 CC The vector encodes soluble or membrane-bound Fc-gamma-RIII, which can be  
 CC used to treat immune thrombocytopenic purpura. Cells transformed with  
 CC the vector, and carrying the protein on the cell surface can be used for  
 CC detecting immune complexes in a sample. The vector is similar to the pCD  
 CC shuttle vector, except that the SV40 promoter has been modified to  
 CC improve expression by the downstream insertion of a portion of the LTR  
 CC from an HIV-I retrovirus. Host cells are mouse cells or COS cells.  
 CC The last 3' end a residue is the start of a polyA sequence.  
 SO Sequence 1990 BP; 592 A; 439 C; 466 G; 493 T;

Query Match 6.2%; Score 15; DB 1; Length 1990;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 AATCTACTATACCA 237  
 DB 1188 AATCTACTATACCA 1174

## RESULT 10

ID N92452/c  
 AC N92452;  
 DT 04-MAY-1990 (first entry)  
 DE cDNA insert of vector pCD(SR alpha)-NL10.  
 KW Vector pCD(SR alpha)-NL10; Fc gamma RIII; ss;  
 KW Immune thrombocytopenic purpura.  
 FH Key Location/Qualifiers  
 FT cds 91..855  
 FT /\*tag= a  
 PN EP-343950-A.  
 PD 29-NOV-1989.  
 PF 24-MAY-1989; 305241.  
 PR 27-MAY-1988; US-199513.  
 PR 19-JAN-1989; US-300039.  
 PA (SCHE) Schering Biotech Corp.  
 PI Peltz GA, Moore KM;  
 DR WPI: 89-349994/48.  
 DR P-PSDB: P91541.  
 PT Soluble and membrane-bound Fc-gamma-RIII - useful for treating immune  
 PT thrombocytopenic purpura, and in assays for immune complexes.



PS Disclosure: Fig. 4A-B; 25pp: English.  
CC The vector encodes soluble or membrane-bound Fc-gamma-RIII, which can be  
CC used to treat immune thrombocytopenic purpura. Cells transformed with  
CC the vector, and carrying the protein on the cell surface can be used for  
CC detecting immune complexes in a sample.  
SO Sequence 2069 BP; 610 A; 445 C; 503 G; 511 T;

Query Match	6.2%;	Score 15;	DB 1;	Length 2069;
Best Local Similarity	100.0%;	Pred. No. 86;		
Matches 15;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 223	AATTCTACTATCA	237		
Db 1259	AATTCTACTATCA	1245		

Sequence	2400 BP	707 A	512 C	468 G	712 T
CC	1				
CC	2				
CC	3				
CC	4				
CC	5				
CC	6				
CC	7				
CC	8				
CC	9				
CC	10				
CC	11				
CC	12				
CC	13				
CC	14				
CC	15				
CC	16				
CC	17				
CC	18				
CC	19				
CC	20				
CC	21				
CC	22				
CC	23				
CC	24				
CC	25				
CC	26				
CC	27				
CC	28				
CC	29				
CC	30				
CC	31				
CC	32				
CC	33				
CC	34				
CC	35				
CC	36				
CC	37				
CC	38				
CC	39				
CC	40				
CC	41				
CC	42				
CC	43				
CC	44				
CC	45				
CC	46				
CC	47				
CC	48				
CC	49				
CC	50				
CC	51				
CC	52				
CC	53				
CC	54				
CC	55				
CC	56				
CC	57				
CC	58				
CC	59				
CC	60				
CC	61				
CC	62				
CC	63				
CC	64				
CC	65				
CC	66				
CC	67				
CC	68				
CC	69				
CC	70				
CC	71				
CC	72				
CC	73				
CC	74				
CC	75				
CC	76				
CC	77				
CC	78				
CC	79				
CC	80				
CC	81				
CC	82				
CC	83				
CC	84				
CC	85				
CC	86				
CC	87				
CC	88				
CC	89				
CC	90				
CC	91				
CC	92				
CC	93				
CC	94				
CC	95				
CC	96				

Query Match	6.2%;	Score 15;	DB 1;	Length 2400;
Best Local Similarity	100.0%;	Pred. No. 86;		
Matches 15;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
0y	121 CAGAAAAAAGTTAG	135		

Db 455 CAGGAAAAAGTTAG 441

RESULT 12  
 086486  
 ID 086486 standard; cDNA; 30 BP.  
 AC 086486;  
 DT 24-NOV-1995 (first entry)  
 DE Primer, e22, for amplification of E2f-1 DNA binding region.  
 KW Primer; E2f-1; E2f-2; transcription; factor; retinoblastoma; pRb;  
 tumour; suppressor; ss.  
 OS Synthetic.  
 PN GB2282814-A.  
 PD 19-APR-1995.  
 PF 07-OCT-1994; 020283.  
 PR 13-OCT-1993; US-136119.  
 PA (MERI) MERCK & CO INC.  
 PI Helmholtz DECK, Ivey-Hoyte M, Olliff AI;  
 DR WPI: 95-141220/19.  
 PT New human transcription factor E2f-2 - involved in cell cycle  
 regulation and useful for drug screening, also related cDNA,  
 PT plasmids and transformed cells.  
 PS Example 1; Page 15; 53pp; English.  
 CC 086484-91 correspond to the E2f-1 primers e20-27 and f120  
 CC respectively. The amplified E2f-1 DNA binding region is  
 CC substantially complementary to the E2f-2 coding region and is  
 CC used as a probe for this region. E2f-2 is involved in cell cycle  
 CC regulation. In particular binding of E2f to the retinoblastoma  
 CC gene product (pRb) causes down regulation of the transcription  
 CC of genes containing the E2f binding site. E2f-2 is useful in  
 CC the study of cell cycle regulation especially in the study  
 CC of pRb and certain viral oncogenes and oncoproteins.  
 SQ Sequence 30 BP; 8 A; 13 C; 4 G; 5 T;

```
Query March          6.28; Score 15; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 95,
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RESULT	13	
T58840_4/c		
Continuation (5 of 6) of T58840 from base 400001 (Mycoplasma genitalium genome.)		
WP	Sequence split into 6 fragments	LOCUS T58840 Accession T58840
WP	Fragment Name	Begin End
WP	T58840_0	1 110000
WP	T58840_1	100001 210000
WP	T58840_2	200001 310000
WP	T58840_3	300001 410000
WP	T58840_4	400001 510000
WP	T58840_5	500001 580073

Query Match	6.2%	Score 15	DB 1	Length 110000
Best Local Similarity	100.0%	Pred. No. 78		
Matches 15	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	224	ATTCTACTATCCAA	238	
db	45810	ATTCTACTATCCAA	45796	

RESULT 14  
T48019/c  
ID T48019 standard; DNA; 27 BP.  
AC T48019;  
DT 29-MAY-1997 (first entry)  
DE Encodes FIV principal immunodominant domain mutant peptide n14

KW Feline immunodeficiency virus; principal immunodominant domain;  
 KW mutant; lentivirus; envelope protein; neutralising antibody;  
 KW vaccine; ss.  
 OS Feline immunodeficiency virus.  
 PN WO9630527-A1.  
 PD 03-OCT-1996.  
 PF 26-MAR-1996; F00449.  
 PR 27-MAR-1995; FR-003566.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PI Pancingo G, Sonigo P;  
 DR WPI: 96-455372/45.  
 P-PSDB: W09332.  
 PT Mutant lenti-virus Env protein principal immuno-dominant domains -  
 PT can produce neutralising antibodies to protect against e.g. FIV or  
 PT HIV, but do not produce facilitating antibodies  
 PS Claim 6: Page 21; 38pp; French.  
 CC The present sequence encodes a specifically claimed example of a novel  
 CC peptide fragment which is a mutant of the wild-type (w.t.) principal  
 CC immunodominant domain (PID) of feline immunodeficiency virus such  
 CC that an Env protein containing the fragment retains the ability to  
 CC produce neutralising antibodies (Abs) but does not produce  
 CC facilitating or deleterious Abs against the w.t. PID. The mutated  
 CC Env protein can be used for producing vaccines against FIV. In  
 CC addition to vaccination, the peptide fragment is useful as a  
 CC diagnostic reagent, specifically to distinguish between Abs produced  
 CC in response to vaccination and those resulting from viral infection.  
 CC Also, the results of anti-FIV vaccination can be monitored  
 CC using the peptide.  
 CC Sequence 27 BP: 8 A; 5 C; 4 G; 10 T;  
 SQ

Query Match 6.2%; Score 15; DB 1; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAATGCTGATGTTTC 15  
 |||||  
 DB 18 AAAATGCTGATGTTTC 4

RESULT 15  
 T60450/c  
 ID T60450 standard; cDNA: 7918 BP.  
 AC T60450;  
 DE 09-JUL-1997 (first entry)  
 DR Melanoma-associated chondroitin sulphate proteoglycan (MCSP) cDNA.  
 KW MCSP; melanoma-associated chondroitin sulphate proteoglycan;  
 KW tumour; sarcoma; glioblastoma; diagnosis; therapy; vaccine; ss.  
 OS Homo sapiens.  
 FH key Location/Qualifiers  
 FT cds 1..6969  
 FT signal\_peptide 1..87  
 FT mat\_peptide 88..6966  
 FT misc\_difference 7434  
 FT /\*tag= d  
 FT /note= "base 7434 is given as n in the  
 FT specification"

FN WO9713855-A1.  
 PD 17-APR-1997.  
 PF 10-OCT-1995; E03988.  
 PR 10-OCT-1995; WO-E03988.  
 PA (NOVS ) NOVARTIS AG.  
 PI Pluschke G, Schmid P;  
 DR WPI: 97-235891/21.  
 P-PSDB: W15566.  
 PT Melanoma-associated chondroitin sulphate proteoglycan - useful for  
 PT prophylaxis or treatment of melanoma-associated tumours  
 PS Claim 6: Page 31-48; 88pp; English.  
 CC A cDNA clone (T60450) codes for melanoma-associated chondroitin  
 CC sulphate proteoglycan (MCSP) (W15566). The complete coding

CC sequence and 3' untranslated region were obt'd. by sequencing small  
 CC overlapping cDNA clones produced by PCR amplification (see also  
 CC T60451-83), and by analysis of cDNA clone lambda M3.1 obt'd.  
 CC from a melanoma cDNA library. MCSP cDNA, esp. sequences excluding  
 CC nucleotides 4867-7898 or 4858-5357, can be used to produce  
 CC recombinant MCSP in host cells for use in vaccines and in methods  
 CC for the treatment and diagnosis of melanoma, sarcoma and  
 CC glioblastoma. Probes (T60462-67) derived from MCSP cDNA can be  
 CC used to identify novel non-human homologues of MCSP and in tumour  
 CC diagnosis, e.g. to localise MCSP mRNA in primary melanomas or  
 CC metastatic lesions.  
 CC Sequence 7918 BP: 1451 A; 2551 C; 2455 G; 1460 T;  
 SQ

Query Match 6.2%; Score 15; DB 1; Length 7918;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 GTGGGCTCATCTACG 187  
 |||||  
 DB 3224 GTGGGCTCATCTACG 3210

Search completed: August 5, 1999, 17:39:47  
 Job time: 6222 sec

---

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:33:17 ; Search time 3268.17 Seconds  
(without alignments)  
233.548 Million cell updates/sec

Title: US-09-049-696-3

Perfect score: 240

Sequence: 1 AAAAACTGATGTTCTGCTT.....GAATTCCTACTATCCAAAG 240

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBank: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_st: \*  
14: gb\_sts: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_v1: \*  
18: em\_fun: \*  
19: em\_hcg: \*  
20: em\_hum1: \*  
21: em\_hum2: \*  
22: em\_in: \*  
23: em\_om: \*  
24: em\_or: \*  
25: em\_ov: \*  
26: em\_pat: \*  
27: em\_ph: \*  
28: em\_pl: \*  
29: em\_ro: \*  
30: em\_sts: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_v1: \*  
34: gb\_hcg1: \*  
35: gb\_hcg2: \*  
36: gb\_in1: \*  
37: gb\_in2: \*  
38: em\_ba1: \*  
39: em\_ba2: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	240	100.0	2826	42	AF127036	Homo sapi

2	189	78.8	3311	11	AF039400	AF039400 Homo sapi
3	151	62.9	35278	11	AF039401	AF039401 Homo sapi
4	20	8.3	2984	3	BT036445	U36445 Bos taurus
5	19	8.3	2937	12	AB017156	AB017156 Mus muscu
6	7.9	2184	5	A30438	A30438 H.sapiens p	
7	19	7.9	2064	9	H01FNRG3	J03143 Human inter
8	9	299	11	HSINEFGAC	U19243 Homo sapien	
9	19	7.9	2088	11	HSU77735	U77735 Human pim-2
10	19	7.9	117198	34	HSU503F13	AL050337 Homo sapi
11	18	7.5	3317	3	AF001261	AF001261 Bos tauru
12	18	7.5	3288	3	AF001262	AF001262 Bos tauru
13	18	7.5	2820	3	AF001263	AF001263 Bos tauru
14	18	7.5	1265	3	AF001264	AF001264 Bos tauru
15	18	7.5	735	5	A39868	A39868 Sequence 2
16	18	7.5	42460	7	SPCC1020	AL033518 S.pombe c
17	18	7.5	152593	8	ATAC002332	AC002332 Arabidops
18	18	7.5	37947	8	SCU39205	U39205 Saccharomyc
19	18	7.5	3259	8	SCU50630	U50630 Saccharomyc
20	17	7.1	122349	1	D90908	D90908 Synchocyst
21	17	7.1	192000	2	AE001438	AE001438 Clostrid
22	17	7.1	10236	2	AE001521	AE001521 Helicobac
23	17	7.1	1560	4	CYIMHCAL1	M37107 C.carpio cl
24	17	7.1	1560	5	E03540	E03540 DNA sequenc
25	17	7.1	82893	7	AB022218	AB022218 Arabidops
26	17	7.1	3007	7	ATHTRPB	M81620 A.thaliana
27	17	7.1	45980	7	ATT24A18	AL035680 Arabidops
28	17	7.1	99345	8	ATAC005623	AC005623 Arabidops
29	17	7.1	91354	8	ATAC006921	AC006921 Arabidops
30	17	7.1	7150	9	AB007931	AB007931 Homo sapi
31	17	7.1	45328	10	AC002128	AC002128 Human DNA
32	17	7.1	165531	10	HSAC002123	AC002123 Human BAC
33	17	7.1	40299	10	HSU65969	U65969 Human Xq28
34	17	7.1	70311	11	AC002390	AC002390 Human DNA
35	17	7.1	141263	11	AC003687	AC003687 Homo sapi
36	17	7.1	163712	11	AC004065	AC004065 Homo sapi
37	17	7.1	175839	11	AC004491	AC004491 Homo sapi
38	17	7.1	127425	11	AC004976	AC004976 Homo sapi
39	17	7.1	119927	11	AC005062	AC005062 Homo sapi
40	17	7.1	146813	11	AF121898	AF121898 Homo sapi
41	17	7.1	3022	12	AF047838	AF047838 Mus muscu
42	17	7.1	3471	12	AF052746	AF052746 Mus muscu
43	17	7.1	712	14	G38088	G38088 RPII-6-112A
44	17	7.1	299427	34	AC006914	AC006914 Caenorhab
45	17	7.1	41009	34	CERY4A7	Z99294 Caenorhabdi

## ALIGNMENTS

RESULT 1  
AF127036  
LOCUS  
DEFINITION  
Homo sapiens calcium-activated chloride channel protein 1 (CACCl)  
AF127036  
mRNA, complete cds.  
ACCESSION  
AF127036  
NID  
AF127036  
VERSION  
AF127036.1 GI:4585468  
KEYWORDS  
AF127036  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 (bases 1 to 2826)  
Agnel,M., Vernal,T. and Culouscou,J.-M.  
Cloning of three human homologs of bovine epithelial chloride  
channel  
Unpublished  
2 (bases 1 to 2826)  
Agnel,M. and Culouscou,J.-M.  
Direct Submission  
Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Rueil-Malmaison 92500, France  
Location/Qualifiers

source	1. .2826	100.0%	Score 240;	DB 427;	Length 2826;
gene	/organism="Homo sapiens"	Best Local Similarity	100.0%;	Pred. No. 2,2e-127;	
CDS	/db_xref="taxon:9606"	Matches 240; Conservative	0; Mismatches	0; Indels	Gaps 0;
	/tissue_type="small intestine; colon"				
	1. .2826				
	/gene="CACCI1"				
	5. .2749				
	/note="bovine epithelial chloride channel homolog"				
	/codon_start=1				
	/product="Calcium-activated chloride channel protein 1"				
	/protein_id="AAD25487.1"				
	/db_xref="GI:94585469"				
	/db_xref="PID:94585469"				
	/translation="MGPEKSSVFILHLHLEALNSLIOLNNNGEIVAIIDPNP EDFLIOOLKDMYQASLIFYEATGKRFYKFNKVAIILPEMKTRADYVRKLEYNKA DYLAESTPENGNDPEPYEQMGNGCEKERSIHLTPDPAAGKLEIYEGOGAGFHEMAH LVKCFEEDINDEKIFYLSNGRIQAVRSAGITGVNVRKQGGSCYTRKTPNKVTGL YKECEPFLQSRQTEKASIMFAQHVDSIVFCTEQNNKEPKNQONCKURSTWEYL RDSDFKRTQPTQPNPTFSLOIQRIVCLVLDSSGSMATGNLRLNQAGQFL LOVLESGWMTGPTDPAAHVQSELIQINSQSDYDLAKRLPAASGSGTSCGLRBA FTVIRKKYPTDGEIVLITDGEDNTIGCFENYEVOSGAIHTVVALGPSAQELELSK MTGLOIYVASDQVONNGLIDAFGLASNGNAVRSQIOLSKGLTLNOSOMMGVTVY DSYGRKOTLITLITTTTTPPOLIMDPSGQKQGGVYNDKRMATQIPCTAKYQWXY SVLASQSTLITLVTSRASNATLPIITYTSKINKTSKRPPELVYTAIHTRGASPIILAA SLALLESYVNGKVTLELNDGAGADATKDDGVSRIFTTIDNGRISVYVRALGGVNA AARRAVIPQSGALYIPGWIENDEIQNNPPPELNMKDQVQKQVCFSTSGGSFVAS DVAPAPIPDLPFGQITDLKAEIHGGSLINTWTAPDDYDGHFAHKYIIRISYILD LRDNFNSLOYNTALIPKEANSEEVLEFKENTFENGDLFAIOAVDKVDLKSII SNIRVSLFIPQTPPEPSPDERSAPCPNHINSITPGIHLIKMKWIGELQLSIA"				
BASE COUNT	875 a 623 c 632 g 656 t				
ORIGIN					
Query Match	100.0%; Score 240; DB 427; Length 2826;				
Best Local Similarity	100.0%; Pred. No. 2,2e-127;				
Matches 240; Conservative	0; Mismatches	0; Indels	Gaps	0;	
QY	1 AAAATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGTAATATGAAACCTACACTG 60				
Db	303 AAAATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGTAATATGAAACCTACACTG 362				
QY	61 ACCAGATGGCAACGTGGAGAGAGGGGGAAGGATCCACCTCACTCTCGTATTCATG 120				
Db	363 ACCAGATGGCAACGTGGAGAGAGGGGGAAGGATCCACCTCACTCTCGTATTCATG 422				
QY	121 CAGGAAAAAAGTTAGCTGATATGACCAACAAGGTAGGGCATTTGCCATGAGTGGGCTC 180				
Db	423 CAGGAAAAAAGTTAGCTGATATGACCAACAAGGTAGGGCATTTGCCATGAGTGGGCTC 482				
QY	181 ATCTACGATGGGAGATTTTGACAGATACAAATAATGATGAGAAATCTATATCAATG 240				
Db	483 ATCTACGATGGGAGATTTTGACAGATACAAATAATGATGAGAAATCTATATCAATG 542				
RESULT 2					
AF039400	AF039400	3311 bp	mRNA	PRI	15-DEC-1998
LOCUS	Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,				
DEFINITION	complete cds.				
ACCESSION	AF039400				
NTD	94009457				
VERSION	AF039400.1	GI:4009457			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
	1 (bases 1 to 3311)				
	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and				
	Pauli,B.U.				
TITLE	Genomic cloning, molecular characterization, and functional				

FEATURES	source
JOURNAL	analysis of human C1cA1, the first human member of the family of Ca2+-activated Cl- channel proteins
MEDLINE	Genomics 54 (2), 200-214 (1998)
REFERENCE	2 (bases 1 to 3311)
AUTHORS	Gruber, A.D., Eblbe, R. and Pauli, B.U.
TITLE	Direct Submission
JOURNAL	Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
source	Location/Qualifiers
gene	1. .3311
CDS	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /map="1p22-p31" /tissue-type="small intestine" 1. .3311 /gene="hC1cA1" 352. 3096 /gene="hC1cA1" /note="transmembrane glycoprotein" /codon_start=1 /product="calcium-dependent chloride channel-1" /protein_id="AAC35428.1" /db_xref="PIR:g4009458" /db_xref="GI:4009458" /translation="MGPFKSSVFLLIHLLEGALSNLIOLNNNGYEGIVADPNP EDELIIQIKDMVTQASLYLEATGRKFRFNVAIIPEYMKIADYVRFKLEYTKA DIVLAESTPGNDEPYTEONGNCEKERSHILPDIAGKRLAGPQKAPEVEMAH LRMVFEDYNNDEKFLYSNRIQAVRSAGITGNNVKKCOGGSCYTRKCFNVTGL YKCECFVLSROTERKASIMEPAOHVDSIVECTEIGNNKKAPNOMOKNLRSTWEYI RDSDEKKTPTMTGPONPNPFTLSIIOLIORIVCLVDKSGSATGNRLNLAQGLPL LQYLETSGWMTPTFDSAAHYQSLLIOLNNSGSDPTLAKPLPAASGCTSGSLREA FTYTRKRPIDGSEIVLLIDGEDNTISGCEYENKQSGAIIHTYALGSAQDELEKSK MTGELQYASDOVONGLIDAFGLSSGNGAVSQRSLQLESKLTJLONNSQMNSTVIV DSYGRKTLFLITWTQPOLILMDPSGKOGGVYVNRKRMAYLIQIPGLAKVTWKY SLQASQTLITVTSRASNTLPIITVTSKNDISKEPSPLVYNAIRQASPIILRA SVTLTISVNGKVTLEILLNGAGADAKDDGYSRFTTYDNGRYSVKRVALGVNVA AARRRVIPQSGALYITGWIENDEIIONNPREPEINKDVOHKQCFSRISGGSFVAS DVPAAPIDLPFPGQITDLKAEIHGSLINLTWAPEDDDHGHAKYIIRISIIID LRDFENSLQVNTTALPKRANSEVEFLFRPENTFENGTDLFLAIDAVDVKLSKI SNIRVSLFIPTPTPTSPDETSADCPNHIINSTIPGIHLIKMKWIGELQLSLIN"
BASE COUNT	1028 a 692 c 742 g 849 t
ORIGIN	
Query Match	78.8%; Score 189; DB 11; Length 3311;
Best Local Similarity	99.6%; Pred. No. 4.3e-98;
Matches 239; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1 AAATGCTGATGTTGTTGCTGCTAGTCTACTCTCCACAGTATATGTAACCTTACATG 60
Ddb	650 AAAATGCTGATGTTGTTGCTGCTAGTCTACTCTCCACAGTATATGTAACCTTACATG 709
QY	61 AACAAATGGGCAACGTGGAAGAGAGGGTGAAGAGATCCACCTCACTCTGATTCATG 120
Ddb	710 AACAAATGGGCAACGTGGAAGAGAGGGTGAAGAGATCCACCTCACTCTGATTCATG 769
QY	121 CAGGAAAAAAGTTACTGTAATATGACCAACAAGTAGGGCATTTTGCATGAGTGGGCTC 180
Ddb	770 CAGGAAAAAAGTTACTGTAATATGACCAACAAGTAGGGCATTTTGCATGAGTGGGCTC 829
QY	181 ATCTACGATGGGAGATATTGACGAGTACAAATATGATGAGAATTTCTATTATCCAAATG 240
Ddb	830 ATCTACGATGGGAGATATTGACGAGTACAAATATGATGAGAATTTCTATTATCCAAATG 889
RESULT	3
AF039401	AF039401 35278 bp DNA PRI 15-DEC-1998
LOCUS	Homo sapiens calcium-dependent chloride channel-1 (hC1cA1) gene,
DEFINITION	complete cds.

ACCESSION AF039401  
MID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 35278)  
Gruber,A.D., Elble,R.C., Ji,H.L., Schreur,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human Clcal, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 2 (bases 1 to 35278)  
99047526  
REFERENCE Gruber,A.D., Elble,R. and Pauli,B.U.  
AUTHORS Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1. 35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1. 1617  
/gene="hCLCA1"  
1. 33522  
/gene="hCLCA1"  
767. 1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587. 1596  
/gene="hCLCA1"  
/note="putative"  
1618. 1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,  
9696..9801,15454..15631,18591..18809,19783..20010,  
22323..22427,24522..24628,26644..26859,27447..27708,  
28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,  
15454..15631,18591..18809,19783..20010,22253..22427,  
24522..24628,26644..26859,27447..27708,28766..28936,  
31840..32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC93429.1"  
/db\_xref="pid:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPFKSVFLLIHLLLEGALNSLIQLNNNGYEGIVAVDPNP  
EDETLLOIKDWYTOGLYFELATGRFAPKNAVLLIPETWKRAIYEPKLETTYNA  
DLVAESTPPGNDERTYEMNGNGEGERHLLPTIAGKKLAIEYPOGKATYHEMAH  
LWGVDFEYNDKERTYLSNGRIDAVRCSAGITETNVKRCQGSCCTKCTKENTKVTGL  
YKSGCEFLVQSROTERKASIMFAQHDHSIVEFCEQHNKEAPKRNQKCNLRSTWEVI  
RSEDEKKTTPMTOPNPPTFSILQIGDSIVCLYDKSGSMATGNRLNLSGQLEFL  
LOTVELGSWAVGTEFDSAHOSELIOINSRDRILAKRLPAAAGSGTSCIGLESLA  
FVIRKRYPTDSEIVLIDGENDTISGGENPKOSGAILIHYALGSPSAOIEELISK  
MTGGLOTYASDOYQNNGLIDAGALSNGAVSQRSIQLESKULTQNSQOMNGYIY  
DSIVGADTLELTWTQPPQIILMDPSGQGGFVVDKNTKAYLIDPGIAGVTKMY  
SLQASQDTLITVTSRASNATLEPIIVTSTKNDTSKFPSPLVVYANIRQAGPILRA  
SVTALIESYNGKTVLELIDNGAGADATDDGYSTRFTTYDNGRYSVKRVALGCVN  
AARRVYPOOSGALYIPGWIENDEIQMNPRIINDVDYOHKQVCSFRTSSGGSFVAS  
DYPNAPIPLPPGQITLDNAELHGSILNTLTATAGDDYDHGTAKRIIRISTSLD

LRDKNESLQVNTALIPKEANSEEFLEKPNENITFENCNDLEFLAIOAVDYDLKSEI  
SNIAVSLFIPPTPPPTPPSPDETSAPCPNHNINSTIPGHIHLKIMMKVIGELQSLIA  
6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 62.9%; Score 151; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 3e-76;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GCTGATGTTCTGTTGCTGATGTTACTCTCCAGTATGATGATACCTTACTGAGCAG 65  
|||||  
DB 6979 GCTGATGTTCTGTTGCTGATGTTACTCTCCAGTATGATGATACCTTACTGAGCAG 7038  
QY 66 ATGGGCAACGTGGAGAGGAGGATGAAGGATCCACCTCCTGATTCATTGACAGA 125  
|||||  
DB 7039 ATGGGCAACGTGGAGAGGAGGATGAAGGATCCACCTCCTGATTCATTGACAGA 7098  
QY 126 AAAAGTTAGCTGAATATGACCAAGGTA 156  
|||||  
DB 7099 AAAAGTTAGCTGAATATGACCAAGGTA 7129  
RESULT 4  
BTU36445 2984 bp mRNA MAM 09-FEB-1996  
LOCUS BTU36445  
DEFINITION Bos taurus calcium-activated chloride channel mRNA, complete cds.  
ACCESSION U36445  
VERSION U36445.1 GI:1184065  
KEYWORDS  
SOURCE cow.  
ORGANISM Bos taurus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
1 (bases 1 to 2984)  
Cunningham, S.A., Awayda, M.S., Buben, J.K., Ismailov, I.I., Ariste, M.P., Bertiev, B.K., Benos, D.J. and Fuller, C.M.  
Cloning of an epithelial chloride channel from bovine trachea  
J. Biol. Chem. 270 (52), 31016-31026 (1995)  
96125078  
2 (bases 1 to 2984)  
Benos, D.J.  
Direct Submission  
Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics, University of Alabama at Birmingham, BHSB 706, Birmingham, AL 35294, USA

FEATURES  
source

1. 2984  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19..2730  
/note="epithelial chloride channel"  
/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="PID:g1184066"  
/db\_xref="GI:1184066"  
/translation="MVPRLTVLFLHLPGMKSSMVNLINGVDGVIYAIPNSVE  
DKLIONTEKEMTEASTVLFHAKTRRVPRNPTSLIPMTWKSSEYIMDKOSYDAE  
VIVANFLKHGDDPTTLOGRGEGKQYHFPNLLTNLEPYSGRAAFHAEKHAH  
RWGIDENYGDPTTISRNTLEATRCSTHTGTNTYKCGGSCITRCRDSQGL  
YEAKCTFIPKSOIARIESIMFQOSLSVTEFCEKTHNEAPNLNKKCNKSGTWDVI  
NMSDFONTSPMTENMPPTQPTFSILKSKORVVCYLDGSGMSSEDRFRNQAQEL  
FLIOIIEKSLGVTGDSVAERNNLTKITDNYENTANLPOEANGSICGLK  
AGFOATIOSOOSTGSEIILLDGENEHSCEIEYKOSGVIHIIAIGPSAKKEET  
LSDMTGCHFFANKDINGLTNAPSRSSSGSITQOTILEKALAITKKVNTETP  
VDSITGNDTFEYVWTIKRPELLLDPKKKTSDFKDKLNIHSARLIPGIAETV  
TWYSLGNHNASPOLITVITTRARSPITPTATAHMSQNTAHPSPVIVAYQVSG  
FLPVLGINVTAIETEDHQVLELMDNGADYVNDGIVSRFTDYDNGNRSILK  
HAEARNNTARLROPNOKALYIPGYENKGIILNPREVEDDLAKIEDFSRLTS  
GGSFTYSGAPGPNHPSVLPNKKIIDEAEKFDHIOLEAPANVLDGKASVYIIRI  
SKSFLLQKDFPDNATVNTSSAKPKAGSDENPFERPEFRLENGENFIAVQALNEA  
NITSEVSNIAQIKRTPEDSPVPLGLTISAINLAIPLAMILIV"

BASE COUNT 1027 a 583 c 561 g 813 t  
ORIGIN

Query Match 8.3%; Score 20; DB 3; Length 2984;  
Best Local Similarity 100.0%; Pred. No. 0.49;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 159 GCATTGTCCATGAGTGGGC 178  
|||||  
Db 472 GCATTGTCCATGAGTGGGC 491

RESULT 5  
LOCUS AB017156 2937 bp mRNA ROD 04-MAR-1999  
DEFINITION Mus musculus gob-5 mRNA, complete cds.  
ACCESSION AB017156  
NID 93721911  
VERSION AB017156.1 GI:3721911  
KEYWORDS GOB-5.  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE Komiya, T., Tanigawa, Y. and Hirohashi, S.  
Cloning and identification of the gene gob-5, which is expressed in  
intestinal goblet cells in mice

JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
MEDLINE 99160866  
REFERENCE 2 (bases 1 to 2937)  
AUTHORS Komiya, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru Komiya, EPRU, JST, Hirohashi Cell Configuration Project, 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan  
(E-mail: tkomi@ncsp.jst.go.jp, Tel: 81-298-47-7563, Fax: 81-298-47-5226)

FEATURES  
source

1..2937  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15..2756  
/gene="gob-5"  
15..2756  
/gene="gob-5"  
/codon\_start=1  
/protein\_id="BA33743.1"  
/db\_xref="PID:d1034712"  
/db\_xref="PID:g3721912"  
/db\_xref="GI:3721912"  
/translation="MESLKSPEVFLIHLLEGVSESLIQLNNGYEGIVYIDHVP  
EDMALIHKDNTVQASPYLFEAGKRRFEKNVAILPESKAKPEYTRPLETEKNA  
DVALSTSPLENDPEYTHIGACGKGIIRHLPDPLAGKLLTOYGQDRTFVEMHA  
FRMVENYNDKFEYLSKGNQVNSAATGKCNHNOAPADNORCLRETFVY  
LYKNCVFPPHNPENKASITMENINSVAETCEKHNNOAPADNORCLRETFVY  
IOESDRTQTPMNPAPREFSLQIGORIVCLYLDKSGSMLNDNDLRNNOARLF  
LQVDEGSWGWATFSAAYVQSELQVNSGADRLDLIKHLPVSGAGSISGGLT  
LFTYIKRYPIDGSEIVLVDGENTLSSCEDLVKQSGAIIHVALPAAKLEQLS  
KMTGQIYSSDOVQNNGLVDAPFALSSGNAALQHSIQLESKRVNLQNNQMGVSI  
VDSVGGKDTLELITWTHPTPIEFMDPSGEONGEILDTTKVAYLVQPGAKYFWK  
YSIOASQOTLITVTSRAASATLPIPIVTPVNNKTEPSPVIVASIRGASPIIR  
ASYATLIESVNGKVTIELDNGAGAKATKNDGVSRPFAFDANGRSVYIMAGLV  
TSDORAPAPRNKRWIDGWLEDEVRNRPRTSYVDKQCLCFRTSSGGSVYAN  
VPAAPSLDPELTQIDLKASIGQNLVLTWAPDDYDHGRASNTIIMSTSYD  
LRDHFNSLQNTGILPKKENSSEIEFEELGNTFGNDIFLAIQAVDKSNLSSEI  
SNIRAVSYFIPQAPPIPEDSTPPCPDISINSTIPGIVHLKIMMKWLGEMQVTLGLH"

BASE COUNT 860 a 718 c 693 g 666 t  
ORIGIN

Query Match 8.3%; Score 20; DB 12; Length 2937;  
Best Local Similarity 100.0%; Pred. No. 0.49;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 162 TTGTCCATGAGTGGGCTCA 181  
|||||  
Db 474 TTGTCCATGAGTGGGCTCA 493

RESULT 6  
LOCUS A30438/c 2184 bp DNA PAT 06-AUG-1996  
DEFINITION H.sapiens PBABLD SacI/Asp7181 fragment.  
ACCESSION A30438  
NID 91567031  
VERSION A30438.1 GI:1567031  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Fountoulakis, M., Garotta, G. and Stueber, D.  
AUTHORS Soluble Interferon-gamma receptors and methods for their production  
TITLE Patent: EP 0393502-A 1 24-OCT-1990;  
JOURNAL F. HOFFMANN-LA ROCHE AG  
FEATURES Location/Qualifiers



```
source 1. .2184
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 688 a 413 c 452 g 631 t
ORIGIN

Query Match 7.9%; Score 19; DB 5; Length 2184;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 TACATATGATGAGAAAT 225
|||||
DB 340 TACATATGATGAGAAAT 322

RESULT 7
LOCUS HUMIFNRG 2064 bp mRNA PRI 08-NOV-1994
DEFINITION Human interferon-gamma receptor mRNA, complete cds.
ACCESSION J03143
NID 9184650
VERSION J03143.1 GI:184650
KEYWORDS interferon receptor.
SOURCE Human lymphoid tissue cell line Raji, CDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2064)
Aguet,M., Dembic,Z. and Meriin,G.
Molecular cloning and expression of the human interferon-gamma
receptor
Cell 55 (2), 273-280 (1988)
JOURNAL MEDLINE 89003065
COMMENT Draft entry, 08-SEP-1988.
by M.Aguet, 08-SEP-1988.
FEATURES
Source location/Qualifiers
1..2064
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6q23-q24"
<1..2064
/note="IFNR-gamma mRNA"
49..1518
/gene="IFNGR1"
49..1518
/gene="IFNGR1"
/note="interferon-gamma receptor"
/codon_start=1
/db_xref="GDB:G00-120-688"
/protein_id="AA52731.1"
/db_xref="PID:9306915"
/db_xref="GI:306915"
/translation="MALFLPLIMVGVSRAEMTADLGSSVPTPTNTYTSYNNP
IYMEQIMPOVPTVEYKNTKYNKSEMDIKAINSHKNCSDHVGSPNSIMTRY
KARVGKESAIKASEFAVCRODKIGPKIDIKREKQIDIDIFPSVNDGDEVD
YDPETCYIRVNVNVRMNGSEIQKILTKEDDCDEIQOLAIPVSSINSQVSAE
GVLHWGVTEKSEKVCITIFNSIKSGSLMIPVALLLFLVLSLIFYIKINPL
KESKISLPSLSIVRSATLETPESKYSVLSYQPSLEKVEEPEEPLSPATVGM
HEDNPGKVEHEELSLTEVYTEENIDPDVPGSHLPTERESSPLSSNSEPGSI
ALNSYRNCSESDHSRNGFPDSSCLSHSSLSDEPPNNKGEIKTSGEILITYIK
APTREGIDPHTVLVDLVDGKESLIGRTIDSEKFS"
BASE COUNT 639 a 383 c 426 g 616 t
ORIGIN 1 bp upstream of EcoRI site; chromosome 6q15-q21.

Query Match 7.9%; Score 19; DB 9; Length 2064;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 TACATATGATGAGAAAT 225
|||||
DB 304 TACATATGATGAGAAAT 286
```

```
RESULT 8
LOCUS HSINFGR3/299 bp DNA PRI 28-MAY-1997
DEFINITION Homo sapiens interferon-gamma receptor alpha chain gene, exon 3.
ACCESSION U19243
NID 9632537
VERSION U19243.1 GI:632537
KEYWORDS 3 of 7
SEGMENT human.
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 5 to 177)
Aguet,M., Dembic,Z. and Meriin,G.
Molecular cloning and expression of the human interferon-gamma
receptor
Cell 55 (2), 273-280 (1988)
JOURNAL MEDLINE 89003065
VERSION 2 (bases 1 to 8; 174 to 188)
Meriin,G., van der Leede,B.J., McKune,K., Knezevic,N.,
Bannwarth,W., Romquin,N., Viegas-Pequignot,E., Kiefer,H., Aguet,M.
and Dembic,Z.
The gene for the ligand binding chain of the human interferon gamma
receptor
Immunogenetics 45 (6), 413-421 (1997)
JOURNAL MEDLINE 97246734
REFERENCE 3 (bases 1 to 299)
Demic,Z.
Direct Submission
Submitted (29-DEC-1994) Zlatko Demic, PRPG, Hoffmann-La Roche AG,
Grenzacherstrasse 124, Basel CH-4002, Switzerland
LOCATION/Qualifiers
1..299
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosTCEP-2.8"
/clone_lib="Cosmid library TCEP-1"
/chromosome="6"
/map="6q23"
/cell_line="P-1"
/cell_type="B-cell"
<1..4
/gene="interferon-gamma receptor alpha chain gene"
/number=2
5..177
/gene="interferon-gamma receptor alpha chain gene"
/number=3
178..>299
/gene="interferon-gamma receptor alpha chain gene"
/number=3
BASE COUNT 94 a 40 c 61 g 104 t
ORIGIN

Query Match 7.9%; Score 19; DB 11; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 TACATATGATGAGAAAT 225
|||||
DB 60 TACATATGATGAGAAAT 42

RESULT 9
LOCUS HSU77735 2088 bp mRNA PRI 20-NOV-1998
DEFINITION Human p1m-2 protooncogene homolog p1m-2h mRNA, complete cds.
ACCESSION U77735
NID 91750275
VERSION U77735.1 GI:1750275
```

KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 2088)  
AUTHORS Baytel,D., Shalom,S., Madgar,I., Weissenberg,R. and Don,J.  
TITLE The human pim-2 proto-oncogene and its testicular expression  
JOURNAL Biochim. Biophys. Acta 1442 (2-3), 274-285 (1998)  
MEDLINE 99023748  
REFERENCE 2 (bases 1 to 2088)  
AUTHORS Baytel,D. and Don,J.  
TITLE Direct Submission  
JOURNAL Submitted (07-NOV-1996) Life Sciences, Bar-Ilan University, Ramat Gan 52900, Israel

FEATURES  
source Location/Qualifiers  
1..2088  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="72m"  
/tissue="testis"  
100..102  
/note="similarity at the amino acid level suggests that, as in the murine pim-2, translation might initiate from a cty codon, which is within a Kozak sequence as an alternative initiation codon"  
186..1190  
/note="similar to murine pim-2 product encoded by GenBank Accession Number L41495; serine/threonine protein kinase"  
/codon\_start=1  
/product="pim-2 protooncogene homolog pim-2h"  
/protein\_id="AAC78506.1"  
/db\_xref="GI:1750276"  
/db\_xref="GI:1750276"  
/translation="MLTRPLQGGPAPPGTPTPPPGKDRFAEAYRLGLPLGKGGFG TYPAGHRTDRLQVATKVI PRNRYVIGMSPLSGSLALVIMTKVAGAGGHPGYRL LMFETQEGFMVLEKRPPLAQDLFDITTEKGPGEPSRCFTGQVYAAIOHCHSGYV HRFIDKENTLIDIRGCARKLIDFGSGLAHDEYTDGDFGRTYSPPEWISRQYHALP ATWVSLGILLYDWVCGDIPFERDQETLEAEHPAPVSPDCALIRCLAPPPSSRPS LFEIILDPMWQTPAEDVTPQPLQRRPCPGVLATLSLMPGLAPNGQSRHPMAMSG "

BASE COUNT 436 a 594 c 528 g 530 t  
ORIGIN

Query Match 7.9%; Score 19; DB 11; Length 2088;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 ATGATGACCCCTACACTGA 61  
|||||  
DB 754 ATGATGACCCCTACACTGA 772

RESULT 10  
HSJ503F13 117198 bp DNA HTG 03-JUN-1999  
LOCUS Homo sapiens chromosome 6 clone J503F13, WORKING DRAFT SEQUENCE, in  
DEFINITION unordered pieces.  
ACCESSION AL050337  
NID 94995633  
VERSION AL050337.4 GI:4995633  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 117198)  
AUTHORS Philimore,B.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT  
On Jun 3, 1999 this sequence version replaced gi:4938320.  
IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: dJ503F13 Contig\_ID: 02079 acc=AL050337 Length: 1642 bp Unfinished: dJ503F13 Contig\_ID: 01588 acc=AL050337 Length: 51717 bp Unfinished: dJ503F13 Contig\_ID: 01205 acc=AL050337 Length: 62239 bp.  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will \* be preserved.

FEATURES  
source Location/Qualifiers  
1..117198  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="J503F13"

BASE COUNT 34604 a 23249 c 23307 g 34429 t 1609 others  
ORIGIN

Query Match 7.9%; Score 19; DB 34; Length 117198;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 TACAAATGATGAGAAAT 225  
|||||  
DB 71250 TACAAATGATGAGAAAT 71268

RESULT 11  
AF001261 3317 bp mRNA MAM 19-NOV-1997  
LOCUS Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.  
ACCESSION AF001261  
NID 92623762  
VERSION AF001261.1 GI:2623762  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 3317)  
AUTHORS Elble,R.C., Widom,J., Gruber,A.D., Abdel-Ghany,M., Levine,R., Goodwin,A. and Pauli,B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3317)  
AUTHORS Elble,R.C., Widom,J., Gruber,A.D., Abdel-Ghany,M., Levine,R., Goodwin,A. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146, Ithaca, NY 14853, USA

FEATURES  
source Location/Qualifiers  
1..3317  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="1"  
/tissue="lung"  
63..2780  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86529.1"  
/db\_xref="PID:92623763"

/db\_xref="GI:2623763"  
/translation="MYCLNVILFLTLHLPGKSSMVNLINNGYGVIAINPSPE  
DEKLIENTKEMVTASTYLEFHATKRREYFRNVSILLPTWKSSEIFIPQESYDAD  
VIVANPYLKGDDPYTLQYRGCEGKGYIHTFPNELLTNFHLTYSNGRVEFMAL  
RMOIPEYNDOPFYISRKNTIATRCSTHITGINVFKKPGSGCITSLCRDSOTG  
LYAKCTFLPKRSQAKESIMFPLSHSVTECEKHTNTEAPNLQNKCKNSWAD  
INMSVDFONTSPMTENMPHTPFLSLSKQRYVCLVLDKSGMSAEDRLFQMNQAAE  
LVLIQVIEGSLVGMVTEFDSVATQNLHTRITDDNYQKTKALPOVANGSTICGL  
KAFQAIHSDOSTSGSEIILLTDGEDNEINSCFEDYKRSGLIHTITAGPSAKELE  
TLSNMGTGRFENKIDITGLTNAFSRISRSISGSIQOALQESKALKITGRKRVNGV  
PVDSTFVNDTFFVWTIQKPEIYLODPKGRKKTSDFKEDKLINSARLOIPGIAET  
GWTYSILNHNASSQMLVTVTRARSPITPVYATAHMSQHTAHPSPIVAYQVSO  
GFLPVGISVIAIETEDHQVTLLEMDNAGRDYKNGDITSRYPEDYKGRYSK  
VHAQANNTARLILROPONKVLVPGYVENGKILNPPEVYDILAKAKIEDPSRLT  
SGGSFTVSGAPPPGNPSVSPSKITDLEAKFEDYIQLSWTAPAGVLDKGRKNSIT  
RISKSEMDROEDNMTLVNTSNLIPKASCKENFEKPEHFEVNGTKYISVOAIN  
EAMLISEVSHIVQAIKFIPLPEDSVHDLGKISEITLAILGLPMISVF"

BASE COUNT 1084 a 676 c 634 g 923 t

Query Match 7.5%; Score 18; DB 3; Length 3317;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 ATTGTCCATGAGTGGGC 178  
|||||

Db 518 ATTGTCCATGAGTGGGC 535

RESULT 12  
AF001262  
LOCUS AF001262 3288 bp mRNA MAM 19-NOV-1997  
DEFINITION Bos taurus clone 2 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
complete cds.  
ACCESSION AF001262  
NID 92623764  
VERSION AF001262.1 GI:2623764  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 3288)  
AUTHORS Eblbe, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
JOURNAL endothelial chloride channel  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 3288)  
TITLE Eblbe, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
JOURNAL Direct Submission  
TITLE Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES  
source Location/Qualifiers  
1..3288  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="2"  
/tissue\_type="lung"  
59..2443  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86530.1"  
/db\_xref="PID:92623765"  
/db\_xref="GI:2623765"  
/translation="AVLCNVILFLTLHLPGKSSMVNLINNGYGVIAINPSPE  
DEKLIENTKEMVTASTYLEFHATKRREYFRNVSILLPTWKSSEIFIPQESYDAD  
VIVANPYLKGDDPYTLQYRGCEGKGYIHTFPNELLTNFHLTYSNGRVEFMAL

RNGIDPEYNDOPFYISRKNTIATRCSTHITGINVFKKPGSGCITSLCRDSOTG  
LEAKCTFLPKRSQAKESIMFPLSHSVTECEKHTNTEAPNLQNKCKNSWAD  
INMSVDFONTSPMTENMPHTPFLSLSKQRYVCLVLDKSGMSAEDRLFQMNQAAE  
LVLIQVIEGSLVGMVTEFDSVATQNLHTRITDDNYQKTKALPOVANGSTICGL  
KAFQAIHSDOSTSGSEIILLTDGEDNEINSCFEDYKRSGLIHTITAGPSAKELE  
TLSNMGTGRFENKIDITGLTNAFSRISRSISGSIQOALQESKALKITGRKRVNGV  
PVDSTFVNDTFFVWTIQKPEIYLODPKGRKKTSDFKEDKLINSARLOIPGIAET  
GWTYSILNHNASSQMLVTVTRARSPITPVYATAHMSQHTAHPSPIVAYQVSO  
GFLPVGISVIAIETEDHQVTLLEMDNAGRDYKNGDITSRYPEDYKGRYSK  
VHAQANNTARLILROPONKVLVPGYVENGKILNPPEVYDILAKAKIEDPSRLT  
SGGSFTVSGAPPPGNPSVSPSKITDLEAKFEDYIQLSWTAPAGVLDKGRKNSIT

BASE COUNT 1043 a 707 c 645 g 893 t

Query Match 7.5%; Score 18; DB 3; Length 3288;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 ATTGTCCATGAGTGGGC 178  
|||||

Db 514 ATTGTCCATGAGTGGGC 531

RESULT 13  
AF001263  
LOCUS AF001263 2820 bp mRNA MAM 19-NOV-1997  
DEFINITION Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
complete cds.  
ACCESSION AF001263  
NID 92623766  
VERSION AF001263.1 GI:2623766  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 2820)  
AUTHORS Eblbe, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
JOURNAL endothelial chloride channel  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 2820)  
TITLE Eblbe, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
JOURNAL Direct Submission  
TITLE Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES  
source Location/Qualifiers  
1..2820  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone="3"  
/tissue\_type="lung"  
/cell\_type="aortic endothelial cells"  
194..2656  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86531.1"  
/db\_xref="PID:92623767"  
/db\_xref="GI:2623767"  
/translation="AVLCNVILFLTLHLPGKSSMVNLINNGYGVIAINPSPE  
DEKLIENTKEMVTASTYLEFHATKRREYFRNVSILLPTWKSSEIFIPQESYDAD  
VIVANPYLKGDDPYTLQYRGCEGKGYIHTFPNELLTNFHLTYSNGRVEFMAL  
RNGIDPEYNDOPFYISRKNTIATRCSTHITGINVFKKPGSGCITSLCRDSOTG  
LEAKCTFLPKRSQAKESIMFPLSHSVTECEKHTNTEAPNLQNKCKNSWAD  
INMSVDFONTSPMTENMPHTPFLSLSKQRYVCLVLDKSGMSAEDRLFQMNQAAE  
LVLIQVIEGSLVGMVTEFDSVATQNLHTRITDDNYQKTKALPOVANGSTICGL  
KAFQAIHSDOSTSGSEIILLTDGEDNEINSCFEDYKRSGLIHTITAGPSAKELE  
TLSNMGTGRFENKIDITGLTNAFSRISRSISGSIQOALQESKALKITGRKRVNGV

BASE COUNT 940 a 565 c 560 g 755 t  
ORIGIN  
PVDSTVGNDFEVVTVTIQKEPIVLQDPKGGKRYKTSDEKDLNIRSLQIPGIAET  
GTWTVSLNNHASSOMLTVTVTRARSPITIPVIAIAMSQHTAIPSPMIVYAVSQ  
GELPIVIGSVIAIETEDGHOVTELEMDNGAGROVDKDIYSRYFTYIGRGVSLK  
VHAORNTARLRINLPONKRVYVPGYENGKIIINPREVEXDKLAKAKIEDFRLT  
SGGSTVSGAPRPNHPSYFPRSKITDLEANKREYIQLSWAPENVLDKNGASSFP  
MSRSHQVAKVLELQLOHOSFO"  
Query Match 7.5%; Score 18; DB 3; Length 2820;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 161 ATTGTCATGAGTGAGC 178  
|||||  
Db 649 ATTGTCATGAGTGAGC 666  
RESULT 14  
AF001264 1265 bp mRNA MAM 19-NOV-1997  
LOCUS Bos taurus clone 4 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.  
ACCESSION AF001264  
NID 92623768  
VERSION AF001264.1 GI:2623768  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 1265)  
Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
Cloning and characterization of Lu-ECAM-1 suggest it is an  
endothelial chloride channel  
Unpublished  
JOURNAL 2 (bases 1 to 1265)  
Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Pauli, B.U.  
Direct Submission  
Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES  
Location/Qualifiers  
1..1265  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cistype="lung"  
/cell\_type="aortic endothelial cells"  
/clone="4"  
109..1137  
/note="endothelial adhesion molecule; chloride channel;  
secreted form"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86532.1"  
/db\_xref="PID:92623769"  
/db\_xref="GI:2623769"  
/translation="MTICLVNVLFLHLPGKSSMVNLINNGYDGIYVIAINPSVPE  
DEKLINIKEMVTEASTYLFHAKRKYFRNNVSIILPMTKSKSEYFIPKOSYDAD  
VIVANPYLKYGDPTLYQVGCGRGKGYIHFTPNFLTNHFIYGRGRVYHEMAHL  
RMGIFDEYNVDQPFYISRNNTLEATRCSTHIGINVFERKPGSGCITSLCRRDSQTG  
LYEAKCTFLPKSQTAKESIMFWSLSVTEFCTEKTHNTAEPNLNKMCKNGKSTWDY  
INNSVDFONTSPMTENMPHTPHFSLKSKQVRVCLVLDKSGSMSEDYLLALIKI  
FKLIGNTI"  
BASE COUNT 440 a 231 c 231 g 363 t  
ORIGIN  
Query Match 7.5%; Score 18; DB 3; Length 1265;  
Best Local Similarity 100.0%; Pred. No. 6.8;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 161 ATTGTCATGAGTGAGC 178  
|||||  
Db 564 ATTGTCATGAGTGAGC 581  
RESULT 15  
A39868/c 735 bp DNA PAT 05-MAR-1997  
LOCUS A39868  
DEFINITION Sequence 2 from Patent EP0614981.  
ACCESSION A39868  
NID 92296098  
VERSION A39868.1 GI:2296098  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 735)  
Demic, Z., Garotta, G. and Gentz, R.H.  
AUTHORS Chimeric human interferon-gamma-receptor/immunoglobulin  
TITLE Polypeptides  
JOURNAL Patent: EP 0614981-A 2 14-SEP-1994;  
HOFFMANN LA ROCHE (CH)  
COMMENT Other publication JP 6319552 941122  
Other publication NZ 280997 951026  
Other publication CA 2114168 940906  
Other publication CN 1094092 941026  
Other publication AU 5647894 940908  
Other publication ZA 9401333 940906.  
FEATURES  
Location/Qualifiers  
1..735  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 223 a 135 c 170 g 207 t  
ORIGIN  
Query Match 7.5%; Score 18; DB 5; Length 735;  
Best Local Similarity 100.0%; Pred. No. 6.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 207 TACATTAATGATGAGAA 224  
|||||  
Db 256 TACATTAATGATGAGAA 239

Search completed: August 6, 1999, 11:33:30  
Job time: 6096 sec



DR P-PSDB: W65088.  
 PT Recombinant DNA encoding Rickettsia surface layer proteins - useful  
 PT for diagnosing typhus and spotted fever and for preparing vaccines  
 PT against them  
 PS Claim 1: Column 11-24; 20pp: English.  
 CC This sequence encodes the Surface layer (S-layer) protein from  
 CC R. prowazekii strain Breinl. This sequence is useful for vaccination  
 CC against typhus and spotted fever rickettsial infection or for diagnosing  
 CC diseases caused by these bacteria. The surface layer protein antigens can  
 CC be produced recombinantly in large quantities  
 SQ Sequence 5319 BP; 1815 A; 766 C; 983 G; 1755 T;  
 14.2%; Score 34; DB 1; Length 5319;  
 Query Match 51.3%; Pred. No. 0.1;  
 Best Local Similarity 51.3%; Pred. No. 0.1;  
 Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
 OY 3 NATGCTGATGTTCTGGTTCAGTCTACTCTCCTCAGGTAATGATGAACTTCACTGAG 62  
 DB 1810 AATAAATGCTCTTGACACTGCTTCTATTCAGTAGAGGAGTCTATATTAATTCACGCT 1869  
 OY 63 CAGATGGCACTCTGGAGAGAGGTCGAAGGATCCACCTCAGTCCGATTTCAATGCA 122  
 DB 1870 GATATAGGTACAGGTGGTGTATATGCTGTACACACACATTACTTACTACAGATGCT 1929  
 OY 123 GGAAAAAGTTCGTAATATGACACCAAGGTA 156  
 DB 1930 TCAAAAATTTAGCACTGATGCGCAATATTA 1963  
 RESULT 3  
 VI9156  
 ID VI9156 standard; cDNA; 1423 BP.  
 AC 28-AUG-1998 (first entry)  
 DE Human XAG growth factor huxag-2 cDNA.  
 KW huxag-2; XAG; growth factor; breast disease; liver disease;  
 KW lung disease; emphysema; wound healing; cancer; diagnosis;  
 KW therapy; human; ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 88..606  
 FT sig\_peptide /\*tag= a  
 FT /\*tag= b  
 FT mat\_peptide 157..603  
 FT /\*tag= c  
 FT WO9807749-A1.  
 PN 26-FEB-1998.  
 PD 22-AUG-1997; U14139.  
 PR 23-AUG-1996; WO-013766.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Dillon PJ, Ebner R, Endress GA, Yu G;  
 DR WPI: 98-169093/15.  
 DR P-PSDB: W37845.  
 PT New isolated human XAG growth factor(s) - used to develop products  
 PT for treating e.g. liver, lung or breast diseases or  
 PT hyperproliferative disorders, e.g. cancer.  
 PS Claim 5: Fig 2A-B; 14pp: English.  
 CC This cDNA clone codes for huxag-2 (see W37845), a member of a novel  
 CC family of human growth factors also including huxag-1 (see W37844)  
 CC and huxag-3 (see W37846). These proteins share homology with the  
 CC XAG protein of *Xenopus laevis*, which is involved in embryogenesis  
 CC and is expressed in adult tissue. huxag-2 cDNA was isolated from a  
 CC cDNA library derived from human microvascular endothelial cells.  
 CC Vectors, host cells, antibodies, and screening methods for agonists  
 CC and antagonists of huxag-2 are provided. huxag polypeptides are  
 CC growth factors and can be used to stimulate proliferation of cells.  
 CC They can be used to stimulate the proliferation and differentiation  
 CC of hepatocytes to alleviate or treat liver diseases and pathologies  
 CC such as fulminant liver failure caused by cirrhosis, liver damage  
 CC caused by viral hepatitis and toxic substances. They can also be  
 CC used to stimulate or promote liver regeneration, e.g. after

CC surgery. They can also be used to prevent and heal damage to the  
 CC lungs caused by various pathological states. They can be used to  
 CC stimulate proliferation and differentiation and promote the repair  
 CC of alveoli and broncholar epithelium to prevent, attenuate, or  
 CC treat acute or chronic lung damage, e.g. emphysema, which results  
 CC in the progressive loss of alveoli, and inhalation injuries, e.g.  
 CC resulting from smoke inhalation and burns, that cause necrosis of  
 CC the broncholar epithelium and alveoli. They can also be used to  
 CC stimulate the proliferation and differentiation of breast tissue  
 CC and could therefore be used to promote healing of breast tissue  
 CC injury due to surgery, trauma or cancer. Antagonists can be used  
 CC to treat hyperproliferative disorders, including cancer, in  
 CC particular hepatocellular carcinoma, osteosarcoma, breast cancer,  
 CC or colon cancer. The products can also be used for detection and  
 CC diagnosis.  
 SQ Sequence 1423 BP; 395 A; 319 C; 315 G; 394 T;  
 13.5%; Score 32.4; DB 1; Length 1423;  
 Query Match 47.9%; Pred. No. 0.21;  
 Best Local Similarity 47.9%; Pred. No. 0.21;  
 Matches 93; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
 OY 47 TGAACCTACAGTACAGAGATGGCACTGTGAGAGAGGTGAAGATCCACTCAC 106  
 DB 285 TGAAGCTTGCAAGACTTAAGCCCAATTTCAGAAATTCAGAACTCTC 344  
 OY 107 TCTGATTTTCATTCAGAAAAAAGTATGCTGAATATGACCAAGTAGGCAATTGT 166  
 DB 345 CCATATATTTTGTATGTTAATCTTGAGAGATGAAGAGAACCAAGATAGATTTCAG 404  
 OY 167 CCATGATGGGCTCATCTACGATGGGAGATTTGACAGTACATATATGAGAAAT 226  
 DB 405 CCTGACGGGGTTATATCCAGATCTTTTCTGATCCAGTGCAGAGGTGATCC 464  
 OY 227 CTAATTCATATG 240  
 DB 465 TGAATCATCATG 478  
 RESULT 4  
 V68522  
 ID V68522 standard; cDNA; 1746 BP.  
 AC 29-JAN-1999 (first entry)  
 DE Nucleotide sequence of the human small secreted protein-1.  
 KW Human small secreted protein-1; SSP-1; hyperproliferative disorder;  
 KW hyperproliferative disorder; cellular communication factor; cytokine;  
 KW cancer; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 591..1109  
 FT /\*tag= a  
 FT /\*product= "human SSP-1"  
 FT WO9846756-A1.  
 PN 22-OCT-1998.  
 PD 15-APR-1998; U07985.  
 PR 16-APR-1997; US-843656.  
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS.  
 PI McCarthy SA;  
 DR WPI: 98-594580/50.  
 DR P-PSDB: W80812.  
 PT New isolated human secreted protein, SSP-1 - used to develop  
 PT products for treating hyperproliferative or hypoproliferative  
 PT disorders, e.g. cancers and fibroproliferative disorders  
 PS Claim 15: Fig 1; 13pp: English.  
 CC This is the nucleotide sequence encoding the human small secreted  
 CC protein-1 (SSP-1) involved in the method of the invention to develop  
 CC products for hyperproliferative or hypoproliferative disorders.  
 CC Based on the fact that SSP-1 is a secreted protein having a small  
 CC molecular size, SSP-1 is likely to be a cellular communication  
 CC factor, such as a cytokine. It is likely to regulate cell  
 CC proliferation, differentiation and/or survival. Alternatively,  
 CC SSP-1 could also interact with an extracellular molecule, e.g. a

CC protein, and be involved in the transport of such molecule, or  
CC : SSP-1 can be a structural protein. The SSP-1 compounds can be used  
CC for modulating growth, differentiation or survival of cells. SSP-1  
CC can be used for treating or preventing a disease associated with an  
CC abnormal cell proliferation, differentiation or survival in a subject,  
CC such as a hyperproliferative or hypoproliferative disease, e.g.  
CC cancers.

Sequence 1746 BP; 445 A; 437 C; 416 G; 448 T;

Query Match 13.5%; Score 32.4; DB 1; Length 1746;  
Best Local Similarity 47.9%; Pred. No. 0.23;  
Matches 93; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Oy 47 TGAACCTCACTAGCAGATGGCACTGTGAGAGAGAGGATGAAAGATCCACCTCAC 106  
Db 788 TGGAGCTTCAAGCTCTAAAGCCCAATTTGCAGATCTACGAAATTTGAGAACTCTC 847  
Oy 107 TCCTGATTCATTCGACGAAAAAGTTAGCTGAATATGACCAAGAGTGGCATTTGT 166  
Db 848 CCATATATTTTGTATGTGTAATCTTGAGGATGAGAGAACCCCAAGATGAAATTTGAG 907  
Oy 167 CCATGATGGGCTCATCTCGATGGGAGATTTGACAGTACATATATGATGAGAAAT 226  
Db 908 CCCTGACGGGGTTATATTCACCAATCTTTTCTGATCCCACTGCGAAGTGCATCC 967  
Oy 227 CTACTATCCAAATG 240  
Db 968 TGAATCATCAATG 981

## RESULT 5

V21209\_05/c  
Continuation (6 of 17) of V21209 from base 500001 (Methanococcus jannaschii circular chr  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 12.6%; Score 30.2; DB 1; Length 110000;  
Best Local Similarity 56.6%; Pred. No. 6.1;  
Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Oy 109 CTGATTTTCATTCGAGAAAAAGTTAGCTGAATATGACACACAGAGGCGATTGTCC 168  
Db 106101 CTGATTTTCATTCGAGAAAAAGTTAGCTGAATATGACACACAGAGGCGATTGTAG 106042

Oy 169 ATGAGTGGGCTCATCTACGATGGGAGATTTGACGAGT 207  
Db 106041 GTCCAGAGCTCCTTAGGGGAGAGTGTTGTATTTGT 106003

## RESULT 6

V21209\_06/c  
Continuation (7 of 17) of V21209 from base 600001 (Methanococcus jannaschii circular chr  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 12.6%; Score 30.2; DB 1; Length 110000;  
Best Local Similarity 56.6%; Pred. No. 6.1;  
Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Oy 109 CTGATTTTCATTCGAGAAAAAGTTAGCTGAATATGACACACAGAGTGGCATTTGTCC 168  
Db 6101 CTGATTTTCATTCGAGAAAAAGTTAGCTGAATATGACACACAGAGTGGCATTTGTAG 6042  
Oy 169 ATGAGTGGGCTCATCTACGATGGGAGATTTGACGAGT 207  
Db 6041 GTCCAGAGCTCCTTAGGGGAGAGTGTTGTATTTGT 6003

## RESULT 7

T12960  
ID T12960 standard; cDNA; 396 BP.  
AC T12960;  
DT 17-NOV-1996 (first entry)  
DE AdunAP4.  
KW AcunAP; HponAP; NamAP; AcenAP; AdunAP; anticoagulant;  
KW nematode-extracted anticoagulant protein; serine protease;  
KW nematode; thrombosis; parasitic worm; ss.  
OS Ancylostoma duodenale.  
FH Key Location/Qualifiers  
FT cds 10..240  
FT /tag= a

PD 25-APR-1996.  
PF 17-OCT-1995; U13231.  
PR 18-OCT-1994; US-326110.  
PR 05-JUN-1995; US-46339.  
PR 05-JUN-1995; US-461965.  
PR 05-JUN-1995; US-463397.  
PR 05-JUN-1995; US-463380.  
PA (CORV-) CORVAS INT INC.  
PI Berghum PM, Ganssems JGJ, Jespers LS, Laroche YR.  
PI Lauwereys MJ, Messens JH, Moyle W, Stanssens PEH.  
PI Vlasuk GP.  
PI WPI: 96-222007/22.  
DR P-PSDB: R91714.  
PT Proteins with anticoagulant and/or serine protease inhibitory  
PT activity - isolated from nematodes and useful to inhibit blood  
PT coagulation  
PS Clam 243; Fig 7D; 243p; English.  
CC Proteins with anticoagulant and/or serine protease inhibitory  
CC activity, isolated from nematodes, are useful to inhibit blood  
CC coagulation. The proteins can be added to blood collection tubes  
CC to prevent or inhibit thrombosis, and may be given alone or in  
CC combination with other therapeutic or in vivo diagnostic agents.  
CC The proteins can serve as immunogens to raise antibodies for use in  
CC the diagnosis and identification of NAP concn. levels in biological  
CC fluids, e.g. to detect mammalian infection with a parasitic worm.  
CC They can also be used as immunogens in prophylactic and therapeutic











---

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 03:37:28 ; Search time 921.96 Seconds

(without alignments)  
60.515 Million cell updates/sec

Title: US-09-049-696-1

Perfect score: 223  
Sequence: 1 GAAATCAGACGAGATGATAC.....ATCTCTGATCTGTTGAAG 223

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50.8	22.8	398	1 X40511	Human secreted pro
2	33	14.8	6924	1 Q67283	Human RPP-beta. H
3	32	14.3	5324	1 N90123	DNA encoding human
4	31.8	14.3	8930	1 V22834	Haemophilus paraga
5	31.4	14.1	1876	1 T48239	Neuronal nicotinic
6	31.4	14.1	1876	1 V12197	Human neuronal nlc
7	31.4	14.1	1590	1 V44687	V274F variant huma
8	30.8	13.8	1783	1 V42186	CD40 ligand gene u
9	30.6	13.7	800	1 Q06346	Synthetic Trichosa
10	30.6	13.7	800	1 Q26500	Trichosanthin synt
11	30.4	13.6	3181	1 N80952	Enkephalinase gene
12	30.4	13.5	3181	1 Q62678	Human enkephalins
13	30.2	13.5	751	1 Q75305	Trichosanthin anti
14	30.2	13.5	7486	1 V22837	Haemophilus paraga
15	29.6	13.3	3150	1 N80433	Bovine interleukin
16	29.6	13.3	3369	1 N81690	Human interleukin
17	29.6	13.0	7032	1 V99801	Canine factor VIII
18	28.6	12.8	1158	1 Q37067	S123-like protein
19	28.4	12.7	1378	1 X13473	Enterococcus faeca
20	28.2	12.6	372	1 Q60002	Human Brain Expres
21	28	12.6	409	1 Q59429	Human Brain Expres
22	27.6	12.4	9822	1 N91774	Entire HIV-2/ST pr
23	27.6	12.4	9473	1 N92768	HIV-2 variant HIV-
24	27.6	12.4	9643	1 N80859	Sequence of entire
25	27.6	12.4	9672	1 Q60616	ROD HIV-2 isolate
26	27.6	12.4	12720	1 Q64211	L. lactis branched
27	27.4	12.3	2306	1 Q05873	Sequence encoding
28	27.4	12.3	2306	1 Q06332	Murine IL-2R beta
29	27.4	12.3	2508	1 T67197	zebrafish retinoid
30	27.4	12.3	984	1 T50759	Sequence of ovine
31	27.4	12.3	999	1 X32820	P. putida alcohol
32	27.2	12.2	2759	1 N80149	Insert of plasmid
33	27.2	12.2	8176	1 V64411	Mouse elf-1 cDNA.
34	27.2	12.2	6991	1 V64412	Mouse elf-2 cDNA.
35	27.2	12.2	3377	1 V64424	Mouse elf cDNA. Ne
36	27.2	12.1	477	1 X20943	POLYNUCLEOTIDE seq
37	26.8	12.0	388	1 Q60440	Human brain Expres
38	26.8	12.0	2612	1 T13943	B. thuringiensis v
39	26.8	12.0	2612	1 T74005	B. cereus VIP3a(b)
40	26.8	12.0	2612	1 V16178	Native DNA sequenc
41	26.8	12.0	9409	1 V52294	Streptococcus pneu
42	26.8	12.0	2612	1 V68063	Bacillus thuringie
43	26.8	12.0	2364	1 V68064	Bacillus thuringie

## ALIGNMENTS

44	26.8	12.0	1946	1 V80304	Human GILR protein
45	26.8	12.0	549	1 V68642	Nucleotide sequenc
RESULT 1					
ID	X40511	standard; cDNA: 398 BP.			
AC	X40511;				
DE	18-JUN-1999 (first entry)				
KM	Human, secreted protein 5', EST SEQ ID No: 111.				
KM	Forensic; gene therapy; chromosome mapping; signal peptide; prostate;				
KM	upstream regulatory sequence; cytokine activity; cell proliferation;				
KM	differentiation; haematopoiesis regulation; tissue growth regulation;				
KM	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;				
KM	thrombolytic; anti-inflammatory; tumour inhibition; ds.				
OS	Homo sapiens.				
PN	W09506350-82.				
PD	11-FEB-1999.				
PF	31-JUL-1998; IB1232.				
PR	01-AUG-1997; US-905144.				
PA	(GENEST) GENEST.				
PI	Duclert A, Dumas Milne Edwards J, Lacroix B;				
DR	WPI: 99-133780/13.				
DR	P-PSDB; Y11789.				
PT	New isolated prostate-derived nucleic acids - used to develop				
PT	products which may have cytokine, immune regulatory, haematopoiesis				
PT	regulating, anti-inflammatory or tumour inhibition activity				
PS	Clain 1; Page 228; 675pp; English.				
CC	X4038 to X40715 represent 5' expressed sequence tags (ESTs) for human				
CC	secreted proteins expressed in prostate, and encode the proteins given in				
CC	Y11716 to Y11993 respectively. The proteins given represent the signal				
CC	peptide and an N-terminal fragment of a secreted protein. The nucleic				
CC	acid sequences can be used for producing secreted human gene products.				
CC	They can also be used to develop products for diagnosis and therapy. The				
CC	proteins obtained may have cytokine activity, cell proliferation and				
CC	differentiation activity, haematopoiesis regulating activity, tissue				
CC	growth regulation activity, reproductive hormone regulating activity,				
CC	chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,				
CC	receptor/ligand activity, anti-inflammatory activity, tumour inhibition				
CC	activity or other activities. The products can be used in forensic, gene				
CC	therapy and chromosome mapping procedures. The sequences can also be used				
CC	for obtaining corresponding promoter sequences. The nucleic acids				
CC	encoding the signal peptides can be used for directing extracellular				
CC	secretion of a polypeptide or the insertion of a polypeptide into a				
CC	membrane, or importing a polypeptide into a cell. 113 T;				
SQ	Sequence 398 BP; 123 A; 86 C; 76 G; 113 T;				
Query Match					
Best local similarity 62.7%; Pred. No. 2; 3e-07;					
Matches 79; Conservative 0; Mismatches 47; Indels 0; Gaps 0;					
QY	98	TTGAGTGAACAAATGCGTATGAGCATGCTTGCATGACCCCAATGTCGAG	157		
DB	220	TACAGCTTCAAGCAATGCGTATGAGCATGCTTGCATGACCCCAATGTCGAG	279		
QY	158	AAATGAACACATCAATCAATTAAGGACATGTCGACCCCAATGTCGATCTG	217		
DB	280	AGATATGAGACCTCATCTCAATCAATTAAGGACATGTCGACCCCAATGTCGAT	339		
QY	218	TTGAGG 223			
DB	340	TTATAG 345			
RESULT 2					
ID	Q67283	standard; cDNA: 6924 BP.			
AC	Q67283;				

Query Match	Best Local Similarity	14.8%	Score 33:	DB 1:	Length 6924:
Matches 84: Conservative	49.7%	Pred. No. 0.4:	85: Indels	0: Gaps	0
53 TCTTGATTCCTTCACCTTCAGAAAGGGCCCTGAGTAATTCACATTCAGCTGAGAACACA	112				
Db 5927 TCTTATTCATCATACATACAGCTTGAGGCCATCTACTTGAAGAACTGAGCTGAGACA	5986				
QY 113 ATGCGTATGAAAGGCAATTCGCTTGCAATCGACCCCAATGTCGCAAGAAATGAAACACTCA	172				
Db 5987 GTCATATTCATGCTCATGTAATGACACTCTCTATTCCTGGACCGACGACCAAAACAAAGC	6046				
QY 173 TTCACCAATTAAGAGACATGTCGACCCGACGACATCTCTGATCTGTTGA	221				
Db 6047 TAGAGAAACAAATTCACAGCTCTGAGCCACATCAATATACAGCAGAGTGA	6095				
RESULT 3					
ID N90123					
AC N90123: standard; DNA: 5524 BP.					
DI 1-NOV-1989 (first entry)					
DE DNA encoding human common acute lymphoblastic leukaemia antigen					
DNA: human common acute lymphoblastic leukaemia antigen;					
KW analgesic; inhibits leukaemia; endopeptidase.					
OS Homo sapiens (Human)					
FT cds	12				
FT /tag= a					
PN WO8905353-A.					
PD 15-JUN-1989.					
PE 01-DEC-1988: U04280.					
PR 04-DEC-1987: US-247915.					
PA (DNA) Dana Farber Cancer Inst.					
PI Reihertz et. Shipb MA, Richardson NE, Ritz J, Sayre PH:					
PT WPI: 89-192699/26.					
DR P-PSDB: P90393.					
PT DNA encoding human common acute lymphoblastic leukaemia antigen					
PR - used for obtaining pure protein for diagnosis and					
PS treatment of medical conditions.					
CC Claim 4: fig 3: 73pp: English.					
CC DNA encoding human common acute lymphoblastic leukaemia					
CC antigen (cHLA) (see corresp. P90393), and its fragments. Binds to					

Query Match	14.3%	Score 32	DB 1	Length 5524
Best Local Similarity	48.9%	Pred. No. 0.78		
Matches 86	Conservative 0	Mismatches 90	Indels 0	Gaps 0
OY	7	ACAGGGAATGTACAGCAATGGGGCCATTAAAGATTGTGTTCATCTTATCTTAC	66	
DB	726	AAAGGCGCTTGACGCAATGTGATTTATGATTTGTGGCCAGATTATTCGTAG	785	
OY	67	CTTACAGAAAGGGCCCTAGTAAATTCATCTCATTCAGCTGACCAACAATGGCTATGAAGC	126	
DB	766	GAAGAAATATTCATCCATGATGAAACACAGCTTGCTTTGAAATGAATAAAGTTATGAA	845	
OY	127	ATTGTGCTGCATGCACCCCAATGTGCCAGAGATGAACACTCATTCACCAAT	182	
DB	846	TTGGAAAAAGAAATGGCAATGCTACGGCTAAACCTGAAGATCGAAATGATCCAAAT	901	
RESULT 4				
ID	V22834/C	standard; DNA; 8930 BP.		
AC	V22834;			
DT	19-AUG-1998	(first entry)		
DE	Haemophilus paragallinarum	antigenic protein encoding DNA #1.		
KV	Haemophilus paragallinarum	antigenic protein; HI antibody; diagnosis;		
OS	Haemophilus paragallinarum	vaccine; chicken infectious coryza; CIC; fowl; ds.		
Key	Location/Qualifiers			
FT	CDS	243..6371		
FT	sig_peptide	/*tag= a		
FT	mat_peptide	/*note= "antigenic protein"		
FT	CDS	243..452		
FT	/*tag= b	453..6368		
FT	/*tag= c	8375..8930		
FT	/*tag= d	/*note= "antigenic protein; no stop codon given"		
PN	MO9812331-A1.			
PD	26-MAR-1998			
PF	12-SEP-1987	J03222.		
PR	19-SEP-1996	JP-271408.		
PA	(KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.			
PI	(KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.			
PR	Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;			
DR	WPI: 98-230318/20.			
DR	P-PSDB: W56319, W56320.			
PT	Antigenic polypeptide from Haemophilus paragallinarum induces HI			
PT	antibody production - and is useful for diagnosis of and preparation			
PT	of vaccines for chicken infectious coryza			
PS	Claim 1; Page 51-69; 108pp; Japanese.			
CC	The present sequence encodes two antigenic proteins derived from			
CC	Haemophilus paragallinarum strain A-221. The antigenic proteins			
CC	stimulate the production of HI antibodies in fowl. The proteins			
CC	and DNA coding for them can be used in the preparation of vaccines			
CC	for the prevention of chicken infectious coryza (CIC). The proteins			
CC	and their antibodies can be used in the diagnosis and treatment of CIC.			
SO	Sequence 8930 BP; 2933 A; 1623 C; 1994 G; 2380 T;			
Query Match	14.3%	Score 31.8	DB 1	Length 8930;
Best Local Similarity	53.7%	Pred. No. 1.1;		
Matches 66	Conservative 0	Mismatches 57	Indels 0	Gaps 0;
OY	41	GTTCTGTTTCACTCTGATCTTCACTCTTGAAGAGGGCCCTGATTAATTCATCTC	100	
DB	5355	GATCTCGGATGATCTTGAAGGCTTCACCACTTAACGACACAGCAAGATCATCAACCAATG	5296	



CC dysfunction, such as cancer, post-herpetic neuralgia, diabetic  
CC neuropathy, osteoarthritis, Alzheimer's or Parkinson's diseases, Kuru,  
CC psychosis and schizophrenia. Probes based on the DNA are used to detect  
CC the DNA in usual hybridisation or amplification tests, while monoclonal  
CC antibodies are used to detect the protein for diagnosis (in vitro or by  
CC in situ immuno-fluorescent assay). Compared with wild-type alpha7 nACHR,  
CC the protein has about 100-fold greater sensitivity to cholinergic  
CC receptor agonists (nicotine or acetylcholine) and response to these  
CC agonists decays more slowly, but the wild-type inward rectification is  
CC retained.  
SQ Sequence 1590 BP; 322 A; 471 C; 449 G; 348 T;

Query Match 14.1%; Score 31.4; DB 1; Length 1590;  
Best Local Similarity 57.7%; Pred. No. 0.74;  
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 57 GATTCTTCACTTCTAGAGGGGCCCTGAGTATTCATCTCAATGCTGAGACACATGG 116  
DB 302 GACTGTTCGTTCCCGAGATGGCCAGATTGGAAACGACATCTTCTATACAGTGC 361  
OY 117 CTATGAGGCAATGTCGTGCATGACCCCAATGTG 153  
DB 362 TGATGAGCGCTTGAAGCCACATTCACACTAAGCG 398

## RESULT 8

VD2186 8  
ID V42186 standard; DNA: 1783 BP.  
AC V42186:  
DT 23-SEP-1998 (first entry)  
DE CD40 ligand gene used in the course of the invention.  
KW CD40 ligand; alteration; immunoreactivity; human cell;  
KW accessory molecule; ligand; AML; gene therapy; treatment; neoplasia;  
KW autoimmune disorder; rheumatoid arthritis; vaccine; ss.  
OS Mus sp.  
PN WO9826061-A2.  
PD 18-JUN-1998.  
PF 08-DEC-1997; U22740.  
PR 01-DEC-1997; US-982272.  
PR 09-DEC-1996; US-032145.  
PA (REGC) UNIV CALIFORNIA.  
PI Cantwell M, Kipps TJ, Sharma S:  
DR WPI; 98-348521/30.  
PT Vectors containing accessory molecule ligand genes - used for  
PT altering immunoreactivity of cells, particularly for treatment of  
PT neoplasia or autoimmune disorders, e.g. rheumatoid arthritis  
PS Disclosure: Page 117; 167pp; English.  
CC The present sequence represents a CD40 ligand gene. The sequence is  
CC used to exemplify the method of the invention. The specification  
CC describes a method for altering the immunoreactivity of human cells  
CC which comprises introducing a gene encoding an accessory molecule  
CC ligand (AML) into the cells so that the AML is expressed on the surface  
CC of the cells. Vectors containing the AML genes can be used in gene  
CC therapy for treating neoplasia or autoimmune disorders such as rheumatoid  
CC arthritis. They can also be used for vaccination to produce immunity  
CC against a virus cell, bacteria, protein, fungus or neoplasia.  
SQ Sequence 1783 BP; 515 A; 411 C; 407 G; 450 T;

Query Match 13.8%; Score 30.8; DB 1; Length 1783;  
Best Local Similarity 48.8%; Pred. No. 1.2;  
Matches 83; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

OY 33 ATTAGAGTTCTGTCTTATCTTACCTTCTAGAGGGCCCTAGTATTC 92  
DB 814 ATATATCTACTGAGCTGATGATGTTCTTACCTTATGAATGCTGTACCATAC 873  
OY 93 ACATCATGCGTGAACAAATGATGAGGCAATGCTGTCATGACCCCATGT 152  
DB 874 AGTACTTCTCTGTCGAACATGCGCCAAAGAAATATAGGACACTAATCATG 933  
OY 153 -3CCAGAGATGAACACTGATTCACAAATTAAGACATGGTGACCCAG 202

DB 934 GTGAGTAGTGCACCTTCTTCTGCTCTTGGAAATACAAACCCAGG 983

## RESULT 9

OY 006346 9  
ID 006346 standard; DNA: 800 BP.  
AC 006346:  
DT 06-FEB-1991 (first entry)  
DE Synthetic Trichosanthin gene.  
KW trichosanthin (TCS); Human Immunodeficiency Virus; inhibition; ss.  
OS Trichosanthes kirilowii.  
FH Key Location/Qualifiers  
FT cds 38..787  
FT /tag- a  
FT /product-mature alpha-TCS

PN WO9012097-A.  
PD 18-OCT-1990.  
PF 04-APR-1990; U01816.  
PR 04-APR-1989; US-333184.  
PA (GENE-) Genelabs Inc.  
PI Platak M, Chow T, Fry K:  
DR WPI; 90-334847/44.  
DR P-PSDB; R07518.  
PT Recombinant trichosanthin protein - with selective inhibitory effect  
PT on viral expression in HIV infected T-cells or monocytes/macrophage.  
PS Example: Fig 11; 102pp; English.  
CC Synthetic gene containing unique restriction sites spaced 20 to 90  
CC bp apart. Introduction of mutations is facilitated by cassette  
CC replacement. The effect of mutations on inhibitory action can be  
CC investigated. For example, a double mutant containing the amino  
CC acid substitutions Glu(160) to Asp and Arg(163) to Lys was found to  
CC be almost 3 logs less active at inhibiting in vitro translation in  
CC rabbit reticulocyte lysate.  
CC See also Q06343-5 and Q06347-Q06351.  
SQ Sequence 800 BP; 207 A; 208 C; 182 G; 203 T;

Query Match 13.7%; Score 30.6; DB 1; Length 800;  
Best Local Similarity 56.4%; Pred. No. 1;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 43 TCTGTTCATCTTGATTTCTTACCTTCTAGAGGGCCCTGAGTATTCATTCAG 102  
DB 134 TATGACATCCCGCTTTCGTAGCTCCCGGCGACCGCTTATGCAATTCAT 193  
OY 103 CTGACAAACATGGCTATGAGGCAATGCTGTCGATCGA 143  
DB 194 CTGACGAATTAGCCGATGAACATATCTCCGTCGCAATCGA 234

## RESULT 10

ID 026500 standard; DNA: 800 BP.  
AC 026500:  
DT 13-JAN-1993 (first entry)  
DE Trichosanthin synthetic gene.  
KW TCS; alpha-trichosanthin; Radix trichosanthin; abortifacient;  
KW ribosome inactivating protein; RIP; HIV-infected human T cells;  
KW human immunodeficiency virus; Trichosanthes kirilowii; ss.  
OS Synthetic.  
PN US5128460-A.  
PD 07-JUL-1992.  
PF 04-APR-1989; 333184.  
PR 04-APR-1989; US-333184.  
PA (GENE-) GENELABS INC.  
PI Chow TP, Fry K, Platak M:  
DR WPI; 92-249485/30.  
DR P-PSDB; R25573.  
PT Nucleic acid encoding trichosanthin protein - which can be used  
PT to inactivate ribosome(s) to inhibit protein synthesis or to  
PT inhibit HIV expression  
PS Example 6; Fig 11; 53pp; English.



CC A synthetic alpha-cricriochosanthin gene was constructed to facilitate  
CC mutational analysis of alpha-TCS in experiments to investigate  
CC structure-function relationships. The synthetic gene contains  
CC unique restriction sites spaced 20-90bp apart. The translation  
CC product of the synthetic gene corresponds to the mature alpha-TCS.  
CC Variants of the protein can be generated by mutagenesis of the  
CC synthetic gene. The variants can then be screened for changes in  
CC ribosome inhibitory and/or HIV-1 inhibitory activities.  
CC See Q26499-Q26505.  
SQ Sequence 800 BP; 207 A; 208 C; 182 G; 203 T;

Query Match	13.7%	Score 30.6;	DB 1;	Length 800;
Best Local Similarity	56.4%;	Pred. No. 1;		
Matches 57;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;

QY	103	CTGAACAAACATGGCTATGAAGCATCTGTCTTCATTCGA	143
Db	194	CTGACGAATTACGCCGATGAACATATCTCCGTGGCAATTCGA	234
QY	43	TCGTGGTTCATCTTGATCTCTCACCCTTCATGAAGGGGCCGAGATATTCATCATTCAG	102
Db	134	TATGACATCCCGCTGTGGCTAGTCTCCCTGCGGGGACGACGCTTATGATTAATTCAT	193

RESULT	11
N80952	
ID	N80952 standard; DNA; 3181 BP.

AC	RefSeq:	
DT	15-NOV-1990 (first entry)	
DE	Enkephalinase gene (human).	
KW	Enkephalinase; immunosuppressant; ss.	
OS	Homo sapiens.	
FT	Key	location/Qualifiers
FT	Cds	3..2249
ET		/*tag =
ET		/product=enkephalinase

PN EP-272928-A.  
PD 29-JUN-1988.  
PF 23-DEC-1987.  
PR 24-DEC-1986; US-946566.  
PR 12-JAN-1987; US-002478.  
PR (GPTH) Genentech Inc.  
PI Malfroy-Camine B, Scofield PR;  
DR WPI; 86-177004/26.  
PT Deoxyribonucleic acid isolates coding for enkephalinase - and prodyn  
PT of enkephalinase, used in therapeutic applications in diagnostic  
PT immunoassay and as immunosuppressants.

The DNA was sequenced from clones isolated from a lambda gt10 library constructed from poly(A<sup>+</sup>) human placental RNA. At posn. 1413 one clone, lambdaDN7 was observed to have a G (coding for Ala.) in clone lambdaDN5 this was found to be an A (Thr). Since the former is identical to the rat amino acid at posn. 465, the latter probably represents an error of the reverse transcriptase synthesis of the mRNA. The DNA can be ligated into an expression vector for prodn. of the enkephalinase protein in a host. The recombinant protein is useful in the treatment of various pathological disorders associated with endogenous peptides such as tachykinins and kinins, eg. acute inflammation, and hyperimmune responses. It is also useful in diagnostic immunoassays, and as immunosuppressants by virtue of their ability to digest chemotactic molecules. This sequence is also published in Ep-279929.

Sequence	3181 BP;	1056 A;	584 C;	653 G;	887 T;	1 other;
5Q						

Query Match	13.68;	Score 30.4;	DB 1;	Length 3181;
Best Local Similarity	48.38;	Pred. No. 2;		
Matches	85;	Conservative	0;	Mismatches 91;
			Indels	0;
			Gaps	0;

7 ACAGGAGATGTACAGCAATGGGCCATTAAAGAGTCTGTGTATCTTGATTTCTTAC 66

Db 711 AAAGAGCTGTACAGCATATGTGGATTATTTATATTCCTGTGGCCAGATTGATTCCTCAG 770

Oy 67 CTTCTAGAAAGGGGCCCGAGTAATTCACCTCATTCACCTGACGTAACCAACATGGCATGAGAGC 126

Db 771 GAAGAAAGATTGCCCATCATCATGTAAGAAACACAGCTGCTTTGGAAATGATTAAGATTATGAA 830

Oy 127 ATTTGCTGTGCATACGACCCCAATGTGGCCAGAGATGAAGAACCTCTTTCAACAAT 182

Db 831 TTGGAAAAAGAAATTTGCCAATGCTACAGCCTAAACCTGAAAGATCAAAATATATCCAT 886

RESULT	12
062678	
ID	062678 standard; cDNA; 3181 BP.
AC	062678; (first entry)
DT	05-JAN-1995
DE	Human enkephalinase cDNA.
KW	Human enkephalinase; kidney; intestine; kinin
KW	kidney brush border neutral proteolase; pituitary; brain; lymph nodes
KW	neutrophils; enkephalins; tachykinins; ss.
OS	Homo sapiens.
EH	key
FT	Location/Qualifiers
FT	3..2249
FT	/*tag- a
FT	21..2246
FT	/*tag- b
FT	1413
FT	/*tag- C
FT	allele
FT	/*note- "Nucleotide is A in a different clone
FT	changing the codon to Thr"

PD EP-596355-A.  
 PN 11-MAY-1994.  
 PF 23-DEC-1987; 117230.  
 PR 24-DEC-1986; US-946566.  
 PR 12-JAN-1987; US-002478.  
 PA (CETH ) GENENTECH INC.  
 PI Malfroy-Camine B, Schofield PR;  
 DR WPI: 94-152785/19.  
 P-PSDB; R52706.  
 PT Method for assay of enkephalinase using dansylated peptide(s) -  
 PT for identification of recombinant enkephalinase  
 PS disclosure; Fig 1; 61pp; English.  
 CC This sequence encodes human enkephalinase.  
 CC This sequence encodes human enkephalinase. Enkephalinase is a  
 CC neutral endopeptidase or kidney brush border neutral proteinase.  
 CC has been isolated from the kidney, intestine, pituitary, brain and  
 CC lymph nodes, and has been detected in many peripheral organs and  
 CC human neutrophils. The distribution of enkephalinase in the brain  
 CC closely parallels the distribution of the enkephalins. Mammalian  
 CC enkephalinase may be used in the treatment of various pathological  
 CC disorders associated with various endogenous peptides, eg.  
 CC tachykinins and kinins.  
 SQ Sequence 3181 BP; 1055 A; 582 C; 657 G; 887 T;

Query Match	13.6%	Score 30.4	DB 1	Length 3181
Best Local Similarity	48.3%	Pred. No. 2		
Matches 85	Conservative	0	Mismatches 91	Indels 0
				Gaps 0

QY 7 ACAGGAGATATACAGCAATGGGGCCATTAAAGTTCTGTCAATCTTGATCTTCAC 66  
 Db 711 AAGAGGCGTTCTACGCATATGTGGATTTTATTTATTTCTGTGGCCGAGATGTCTCTGAG 770  
 QY 67 CTCTAGAGGGGCCCGATTAATTCATCTATCACTGACGACACCAATGGCATATGAAGC 126  
 Db 771 GAAGAAAGATTGCCCATCGATGAAGAAACCAGCTTGCTTGGAAATGATATAAGTTATGAA 830  
 QY 127 ATGTGCGTTGCATCGACCCCAATGTGCCAGAGATGAAGAAACATCATTCACAAAT 182  
 Db 831 TTGGAAAAAGAAATTTGCCAATGCTACGCGTAAACCTGMAAGATGGAATATATCCAA 886

RESULT 13  
Q75305

ID	Query Match	Best Local Similarity	Matches	Score	DB 1	Length	DB 2	Score	DB 3	Length	DB 4	Score	DB 5	Length	DB 6	Score	DB 7	Length	DB 8	Score	DB 9	Length	DB 10	Score	DB 11	Length	DB 12	Score	DB 13	Length	DB 14	Score	DB 15	Length	DB 16	Score	DB 17	Length	DB 18	Score	DB 19	Length	DB 20	Score	DB 21	Length	DB 22	Score	DB 23	Length	DB 24	Score	DB 25	Length	DB 26	Score	DB 27	Length	DB 28	Score	DB 29	Length	DB 30	Score	DB 31	Length	DB 32	Score	DB 33	Length	DB 34	Score	DB 35	Length	DB 36	Score	DB 37	Length	DB 38	Score	DB 39	Length	DB 40	Score	DB 41	Length	DB 42	Score	DB 43	Length	DB 44	Score	DB 45	Length	DB 46	Score	DB 47	Length	DB 48	Score	DB 49	Length	DB 50	Score	DB 51	Length	DB 52	Score	DB 53	Length	DB 54	Score	DB 55	Length	DB 56	Score	DB 57	Length	DB 58	Score	DB 59	Length	DB 60	Score	DB 61	Length	DB 62	Score	DB 63	Length	DB 64	Score	DB 65	Length	DB 66	Score	DB 67	Length	DB 68	Score	DB 69	Length	DB 70	Score	DB 71	Length	DB 72	Score	DB 73	Length	DB 74	Score	DB 75	Length	DB 76	Score	DB 77	Length	DB 78	Score	DB 79	Length	DB 80	Score	DB 81	Length	DB 82	Score	DB 83	Length	DB 84	Score	DB 85	Length	DB 86	Score	DB 87	Length	DB 88	Score	DB 89	Length	DB 90	Score	DB 91	Length	DB 92	Score	DB 93	Length	DB 94	Score	DB 95	Length	DB 96	Score	DB 97	Length	DB 98	Score	DB 99	Length	DB 100	Score	DB 101	Length	DB 102	Score	DB 103	Length	DB 104	Score	DB 105	Length	DB 106	Score	DB 107	Length	DB 108	Score	DB 109	Length	DB 110	Score	DB 111	Length	DB 112	Score	DB 113	Length	DB 114	Score	DB 115	Length	DB 116	Score	DB 117	Length	DB 118	Score	DB 119	Length	DB 120	Score	DB 121	Length	DB 122	Score	DB 123	Length	DB 124	Score	DB 125	Length	DB 126	Score	DB 127	Length	DB 128	Score	DB 129	Length	DB 130	Score	DB 131	Length	DB 132	Score	DB 133	Length	DB 134	Score	DB 135	Length	DB 136	Score	DB 137	Length	DB 138	Score	DB 139	Length	DB 140	Score	DB 141	Length	DB 142	Score	DB 143	Length	DB 144	Score	DB 145	Length	DB 146	Score	DB 147	Length	DB 148	Score	DB 149	Length	DB 150	Score	DB 151	Length	DB 152	Score	DB 153	Length	DB 154	Score	DB 155	Length	DB 156	Score	DB 157	Length	DB 158	Score	DB 159	Length	DB 160	Score	DB 161	Length	DB 162	Score	DB 163	Length	DB 164	Score	DB 165	Length	DB 166	Score	DB 167	Length	DB 168	Score	DB 169	Length	DB 170	Score	DB 171	Length	DB 172	Score	DB 173	Length	DB 174	Score	DB 175	Length	DB 176	Score	DB 177	Length	DB 178	Score	DB 179	Length	DB 180	Score	DB 181	Length	DB 182	Score	DB 183	Length	DB 184	Score	DB 185	Length	DB 186	Score	DB 187	Length	DB 188	Score	DB 189	Length	DB 190	Score	DB 191	Length	DB 192	Score	DB 193	Length	DB 194	Score	DB 195	Length	DB 196	Score	DB 197	Length	DB 198	Score	DB 199	Length	DB 200	Score	DB 201	Length	DB 202	Score	DB 203	Length	DB 204	Score	DB 205	Length	DB 206	Score	DB 207	Length	DB 208	Score	DB 209	Length	DB 210	Score	DB 211	Length	DB 212	Score	DB 213	Length	DB 214	Score	DB 215	Length	DB 216	Score	DB 217	Length	DB 218	Score	DB 219	Length	DB 220	Score	DB 221	Length	DB 222	Score	DB 223	Length	DB 224	Score	DB 225	Length	DB 226	Score	DB 227	Length	DB 228	Score	DB 229	Length	DB 230	Score	DB 231	Length	DB 232	Score	DB 233	Length	DB 234	Score	DB 235	Length	DB 236	Score	DB 237	Length	DB 238	Score	DB 239	Length	DB 240	Score	DB 241	Length	DB 242	Score	DB 243	Length	DB 244	Score	DB 245	Length	DB 246	Score	DB 247	Length	DB 248	Score	DB 249	Length	DB 250	Score	DB 251	Length	DB 252	Score	DB 253	Length	DB 254	Score	DB 255	Length	DB 256	Score	DB 257	Length	DB 258	Score	DB 259	Length	DB 260	Score
----	-------------	-----------------------	---------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------

```

stimulates the production of HI antibodies in fowl. The protein
CC and DNA coding for it can be used in the preparation of vaccines
CC for the prevention of chicken infectious coryza (CIC). The protein
CC and its nucleotides can be used in the diagnosis and treatment of CIC.
SQ Sequence 7486 BP; 2435 A; 1332 C; 1684 G; 2035 T;

Query Match      13.5%; Score 30.2; DB 1; Length 7486;
Best Local Similarity 52.8%; Pred. No. 3.3;
Matches 65; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 41 GTTCGTGTCATCTGTGATTCCTTACCTTGTAGAAAGGGCCCTGATATTCATCTTC 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5951 GATCCCGCTGGATCTTTGGCTCCACCTTAACGGCAGCAGCAAGATACCAACCAATG 5892

QY 101 AGCTAACACAAATGGCATGAGCATTTGTCGTGCAATGAGCCCAATGTCAGAAAG 160
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5891 CCGTTTATCTCTGCAATGCTGCTGTGTGCGGCGCAATATCCCGCTGTACATTAC 5832

QY 161 ATG 163
    ||
DB 5831 CTG 5829

RESULT 15
N80433/c
ID N80433 standard; cDNA; 1750 BP.
AC N80433;
DT 05-NOV-1990 (first entry)
DE Bovine interleukin-1 beta clone.
KW Interleukin 1 beta; alveolar macrophages; B-lymphocytes; ss.
OS Bovls.
FH Key
FT signal_peptide 74..412
FT mat_peptide 413..871
FT /tag= a
FT /tag= b
FT /product=bil-1beta

WO8806483-A.
22-SEP-1988.
PF 13-FEB-1988; U00406.
PR 13-MAR-1987; US-025462.
PA (IMMU-) Immunex Corp.
PI Cerreti DP, Davis BS, Malleswesi CR;
DR WPI. 88-285368/40.
P-P8DB; P80788.
PT Mammalian, esp. bovine, homologues of human interleukin-1 beta - used for
PT potentiating immune response to vaccine antigens and for wound healing.
PS Disclosure; P; English.
CC The clone, prepd. from RNA isolated from bovine alveolar macrophage, can
CC be used to construct an expression vector for prodn. of recombinant bil-1
CC beta. Therapeutic compns. contg. bil-1 proteins or active homologues
CC can be used to potentiate antibody response to vaccines and also to
CC promote rapid epidermal wound healing.
SQ Sequence 1750 BP; 497 A; 443 C; 397 G; 413 T;

Query Match      13.3%; Score 29.6; DB 1; Length 1750;
Best Local Similarity 56.0%; Pred. No. 2.9;
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 120 TGAAGCATGTGTCGTGCAATGACCCCAATGTGCCAGAGATGAACACTTCATCA 179
    |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 257 TGAAGTTTGTGTAGAACTGTGAGAAATCTGCAGCTGATGTTCCATCTCCATGG 198

QY 180 AATTAAGCATGTGATGACCCAGCATCTCTATCTGTT 219
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197 AACCGAGTCCAGGTGTGATGATCAGCTTCATCTGTT 158
  
```

Streptococcus pneumoniae, for treatment or prevention of infection e.g. pneumonia, otitis media or meningitis. Probes based on the nucleic acid are used to detect Streptococcus infection (by usual hybridisation or amplification methods), also for isolating Streptococcus genes or their allelic variants. The protein can be used similarly to detect specific antibodies in standard immunoassays, especially for diagnosing or monitoring infections. Antibodies which bind the protein are used to detect corresponding antigens, to purify the protein and for passive immunisation (optionally coupled to a toxin). Vaccines are administered, e.g. by injection, orally or through the skin, typically at 0.01-1000 (especially 10-300)  $\mu\text{g/ml}$  per dose.

Sequence 352 BP; 107 A; 65 C; 80 G; 100 T;

[illegible][illegible]

RESULT	15
V43012	
ID	V43012 standard; DNA; 1660 BP.
AC	V43012:
DT	09-NOV-1998 (first entry)
DE	<i>Streptococcus pneumoniae</i> polypeptide coding region.
KW	Polypeptide; ORF; open reading frame; infection; bacterial;
KN	<i>streptococcal</i> ; <i>bacteremia</i> ; <i>diagnosis</i> ; <i>prophylaxis</i> ; <i>ds</i> .
OS	<i>Streptococcus pneumoniae</i> .
EH	key
FT	Location/Qualifiers
CDS	complement(460..740)

RESULT	14
ID	V52162/c
AC	V52162 standard; DNA; 10254 BP.
DT	V52162; 23-OCT-1998 (first entry)
DE	<i>Streptococcus pneumoniae</i> genome fragment SEQ ID NO:29.
KW	<i>Streptococcus pneumoniae</i> ; <i>S. pneumoniae</i> ; genome; diagnosis; assay;
OS	computer readable medium; vaccine; pharmaceutical composition; ds.
KM	<i>Streptococcus pneumoniae</i> .
PN	MO9818931-A2.
PD	07-MAY-1998
PF	30-OCT-1997; U19588.
PR	31-OCT-1996; US-029360.
PA	(HUMA-) HUMAN GENOME SCI INC.
PI	Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI	Kunsch CA, Rosen CA;
DR	WPI: 98-272225/24.
PT	Computer-readable medium with recorded <i>Streptococcus pneumoniae</i>
PT	polynucleotide sequences - useful in diagnostic kits and assays, and
PT	pharmaceutical compositions and vaccines for <i>Streptococcus</i>
PT	<i>pneumoniae</i>

PN MO98823631-A1.  
 PD 04-JUN-1998.  
 PE 24-NOV-1997: U21976.  
 PF 27-NOV-1996: US-031879.  
 PG (SMIK ) SMITHKLINE BEECHAM CORP.  
 PH (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Ionetto MA, Nicholas RO,  
 PJ Reid RH, Zarfos PN.  
 PK WPI: 98-322654/28.  
 PL WPI: 98-322654/28.  
 PM P-PSDB: W62737.  
 PN Streptococcus pneumoniae polynucleotides - useful for developing  
 PT products for diagnosis, prevention and treatment of infections e.g.  
 PU pneumonia, bacteraemia, meningitis or endocarditis  
 PS Claim 1; Page 131-132; 181pp; English.  
 QC The sequence is that of a Streptococcal polypeptide coding region.  
 CC The polypeptide can potentially be used for the diagnosis and  
 CC prevention of bacterial infections, especially SP infection.  
 CC It may be used for the treatment of diseases such as otitis media,  
 CC conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural  
 CC empyema, endocarditis or infection of the cerebrospinal fluid.  
 QD Sequence 1660 BP; 469 A; 392 C; 323 G; 476 T;

CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (VS2134 to VS2524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (VS2134 to VS2524) are genomic fragments from *Streptococcus*  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the *S. pneumoniae*  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridise to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the *S. pneumoniae* genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC *S. pneumoniae* genome of commercial importance, or expression modulating  
CC fragments of the *S. pneumoniae* genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for *S. pneumoniae*.  
CC Sequence 10254 BP: 2974 A; 2014 C; 2354 G; 2911 T;  
CC

```

Query Match          9.4%:  Score 27.2:  DB 1:  Length 1660:
Best Local Similarity 52.2%:  Pred. No. 13:
Matches 59:  Conservative 0:  Mismatches 54:  Indels 0:  Gaps 0:

OY      33  TGCCTGACCGATGGGGAACAACAACACTATTAAGTGGGTGCTTTAACGAGGTCAAACAAGTN 92
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      1359  TACTGGTTGTTGATGACACGCCAAATAATAGGTATCTGTTTCAAGAGGCATCTGGTCCA 1418

OY      93  GTGCCATCATCCACACAGACTCGCTTGGGGGCCCTCTGCAGCTCAAGACTAGAG 145
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      1419  AATTCAAATCTCAATGGGGAATTTTCCCTGCTCTCGGAACCCAAATATTCAG 1471

Search completed: August 6, 1999, 04:02:57
Job time: 1529 sec

```

Db 8420 AGCGTATTGCA 8408

# RESULT 11

ID T13347

AC T13347 standard; cDNA to mRNA; 4062 BP.

DT 26-JUL-1996 (first entry)

DE C3G protein gene.

KW C3G; ras protein guanine nucleotide exchange factor; diagnosis;

KW treatment; malignant tumour; activation; ss.

OS Homo sapiens.

Key Location/Qualifiers

FT cds 123..3356

FT /tag= "a"

FT /note= "C3G protein"

FN J07051073-A.

PD 28-FEB-1995.

PF 13-JUN-1994; 130699.

PR 11-JUN-1993; JP-140806.

PA (SHKJ) SHINGIJUTSU JIGODAN.

DR WPI: 95-127357/17.

P-PSDB: R91597.

PT Ras protein guanine nucleotide exchange factor C3G gene - useful for

PT diagnosis and treatment of malignant tumours associated with ras

PS Claim 2; Page 6; 9pp; Japanese.

CC The present sequence encodes C3G protein which is a ras protein guanine

CC nucleotide exchange factor. The gene and protein (see R91597) are useful

CC for diagnosis or treatment of malignant tumours associated with

CC activation of the ras gene.

SO Sequence 4062 BP; 937 A; 1194 C; 1147 G; 784 T;

Query Match 9.4%; Score 27.2; DB 1; Length 4062;

Best Local Similarity 51.2%; Pred. No. 19;

Matches 63; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 77 GAGGTCAAAAGTNGTCATCATCCACACAGTGGGCGCCCTGACGTCAA 136

Db 1757 GAGGAAAAACAAACATGCTGCGCTCATGCTGAGGAGCTACTCGGAGCCGA 1816

OY 137 GAACAGAGAGAGCTGTCAAAATGACAGAGGTTTACAGACATATCTTCAGTCAATT 196

Db 1817 GCCCTATATGTTCTACAGAGCCAGACAGACATCTACACAGAGAGAGAGT 1876

OY 197 C 197

Db 1877 C 1877

RESULT 12

ID T28527

AC T28527 standard; DNA; 631 BP.

DT 01-APR-1997 (first entry)

DE S. pneumoniae detection probe #1.

KW Detection; probe; amplification primer; bacterial pathogen; pneumonia;

KW Escherichia coli; Klebsiella pneumoniae; Pseudomonas aeruginosa;

KW Streptococcus pneumoniae; Staphylococcus aureus;

KW Staphylococcus epidermidis; Enterococcus faecalis; respiratory tract;

KW Haemophilus influenzae; Moraxella catarrhalis; urinary tract;

KW infection; intra-abdominal infection; skin infection;

KW bacterial resistance; beta-lactam antibiotic; ds.

OS Synthetic.

PN WO9608582-A2.

PD 21-MAR-1996.

PF 12-SEP-1995; CA0528.

PR 12-SEP-1994; US-304732.

PA (BERG/) BERGERON M G.

PA (OUEL/) OUELLETTE M.

PA (ROY/) ROY P H.

PI Bergeron MG, Ouellette M, Roy PH;

DR WPI: 96-179953/18.

PT Method for the detection of bacterial species using probes and

PT primers - allows detection and quantification of antibiotic

PS resistant bacteria in patients, the environment and food

PS Claim 52; Page 88; 216pp; English.

CC The sequences given in T28527/32 represent probes which were used in

CC the method of the invention for the detection of S. pneumoniae in a

CC sample. The method comprises using probes and/or amplification primers

CC which are specific, ubiquitous and sensitive for determining the presence

CC and/or amount of nucleic acids from selected bacterial species in any

CC sample, where the bacterial nucleic acid comprises a selected target

CC region hybridisable with the probes or primers. The method comprises

CC contacting the sample with the probes or primers. The method comprises

CC the presence and/or amount of hybridised primers or amplification products

CC as and indication of the presence and/or amount of the bacterial

CC species. This method may be used to detect commonly encountered

CC bacterial pathogens, e.g. Escherichia coli, Klebsiella pneumoniae,

CC Pseudomonas aeruginosa, Proteus mirabilis, Streptococcus pneumoniae,

CC Staphylococcus aureus, Staphylococcus epidermidis, Enterococcus

CC faecalis, Staphylococcus saprophyticus, Streptococcus pyogenes,

CC Haemophilus influenzae and Moraxella catarrhalis. These bacterial

CC species are associated with approx. 90% of urinary tract infections and

CC with a high percentage of other severe infections including

CC septicemia, meningitis, pneumonia, intra-abdominal infections, skin

CC infections and other severe respiratory tract infections. The method

CC may also be used to evaluate a bacterial resistance to beta-lactam

CC antibiotics.

SO Sequence 631 BP; 174 A; 154 C; 109 G; 194 T;

Query Match 9.4%; Score 27.2; DB 1; Length 631;

Best Local Similarity 52.2%; Pred. No. 8.3;

Matches 59; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 33 TGCTACGAGATGGGAGACACACTATAGTGGTCTTAAAGAGTCAAAAGTGN 92

Db 182 TACGTGTTGTGATGACGTGCCAAATATAGTATCTCTTCAAGAGTCATCTGGTCCA 241

OY 93 GTGCCATCATCCACACAGTGGGCGCCCTGAGTCAAGACTAGAG 145

Db 242 AATTCACATCATGAGGGAATTTCTGCTCTGGAACCAAAATATTACAG 294

RESULT 13

ID V27387/C

AC V27387 standard; DNA; 352 BP.

DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae Sp0076 nucleotide.

KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;

KW detection; pneumonia; otitis media; meningitis; ss.

OS Streptococcus pneumoniae.

Key Location/Qualifiers

FT CDS 2..352

FT /tag= "a"

FT /product= "Sp0076"

FT /note= "no stop codon given"

FN WO9818930-A2.

PD 07-MAY-1998.

PF 30-OCT-1997; U19422.

PR 31-OCT-1996; US-029960.

PA (HUNA-) HUMAN GENOME SCI INC.

PI Choi GH, Hromockyj J A, Johnson LS, Kunsch CA;

DR WPI: 98-272224/24.

P-PSDB: W61201.

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus

PT pneumoniae - or their epitope-containing fragments, useful in

PS protective or therapeutic vaccines, and for diagnosis

PS Claim 1; Page 76; 118pp; English.

CC The present sequence encodes a protein from Streptococcus pneumoniae.

CC The nucleic acid sequence encoding the streptococcus pneumoniae protein

CC can be useful in vaccines for inducing protective antibodies against

Query Match 9.6%; Score 27.8; DB 1; Length 302;  
Best Local Similarity 47.7%; Pred. No. 3.8;  
Matches 74; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY 94 TGCATCATCCACAGAGCTTGGGCGCTTCGACGCTCAAGAACTAGAGAGCTGTC 153  
DB 65 TGAGACCTCCATACAGTTTCCATGATGATGAGGATGCGACGCTTCCTCACCCTG 124  
OY 154 CAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTACAGAAATGGCTCAT 213  
DB 125 AGTACAGCTGGGATGATGGGCTCCGCGCATATACAGGAAACATTCACACCGT 184  
OY 214 TGATGCTTTGGGCGCTTCATCAAGAAATGAG 248  
DB 185 TGATACATGTGACCCCTTGACAGAGGATGAG 219

## RESULT 8

V21209\_03/c  
Continuation (4 of 17) of V21209 from base 300001 (Methanococcus jannaschii circular chr  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

Fragment Name Begin End  
WP V21209\_00 1 110000  
WP V21209\_01 100001 210000  
WP V21209\_02 200001 310000  
WP V21209\_03 300001 410000  
WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000  
WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000  
WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976

Query Match 9.6%; Score 27.8; DB 1; Length 110000;  
Best Local Similarity 53.2%; Pred. No. 51;  
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 130 AGCTCAGAACTAGAGAGCTTCACAAATGACAGAGGTTTACAGACATATGCTTCAGA 189  
DB 107997 ATCATATAAATAAAGAAATGACAAATTCCTTGGAATTTGTGATGCTCGCTTCACT 107938  
OY 190 TCAAGTTCAAGAACTAGGCTTATGATGCTTTGGGCGCTTCATCAGG 240  
DB 107937 CCGATTGCCAAATATGATGATGATGATGATGATGATGATGATGATGATGATG 107887

## RESULT 9

V21209\_04/c  
Continuation (5 of 17) of V21209 from base 400001 (Methanococcus jannaschii circular chr  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

Fragment Name Begin End  
WP V21209\_00 1 110000  
WP V21209\_01 100001 210000  
WP V21209\_02 200001 310000  
WP V21209\_03 300001 410000  
WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000  
WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000

WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976

Query Match 9.6%; Score 27.8; DB 1; Length 110000;  
Best Local Similarity 53.2%; Pred. No. 51;  
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 130 AGCTCAGAACTAGAGAGCTTCACAAATGACAGAGGTTTACAGACATATGCTTCAGA 189  
DB 7997 ATCATATAAATAAAGAAATGACAAATTCCTTGGAATTTGTGATGCTCGCTTCACT 7938  
OY 190 TCAAGTTCAAGAACTAGGCTTATGATGCTTTGGGCGCTTCATCAGG 240  
DB 7937 CCGATTGCCAAATATGATGATGATGATGATGATGATGATGATGATGATGATG 7887

## RESULT 10

X13032/c  
X13032 standard; DNA; 8726 BP.

AC X13032;  
DT 19-MAR-1999 (first entry)  
DE Enterococcus faecalis genome contig SEQ ID NO:95.  
KM Enterococcus faecalis; contig; detection; Enterococcal infection;  
OS Enterococcus faecalis.  
PN MO9850555-A2.  
PD 12-NOV-1998.  
PF 04-MAY-1998; D08985.  
PR 14-NOV-1997; US-066009.  
PR 06-MAY-1997; US-044031.  
PR 16-MAY-1997; US-046655.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Dillon PJ, Kunsch CA;  
DR WPI: 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.

PS Claim 1; Page 626-631; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13191 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
CC Sequence 8726 BP; 2834 A; 1460 C; 1912 G; 2509 T;

Query Match 9.5%; Score 27.4; DB 1; Length 8726;  
Best Local Similarity 50.4%; Pred. No. 23;  
Matches 67; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 138 AACTAGAGAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTC 197  
DB 8540 AACGAAATATCATTAATAAAGTTACAGAAATATGCAATCACTATCTGATAATATCG 8481  
OY 198 AGAACATGGGCTCATGATGCTTTGGGCGCTTCATCAGAAATGAGAGCTTCCTC 257  
DB 8480 CTAAATATCTCTATAGACATTTGGGTGATGCAAAAGCAAAAGCAAAATGATATC 8421  
OY 258 AGCGTCCATCCA 270  
|||||

AC T69307;  
DE 26-FEB-1998 (first entry)  
DE Murine metastatic nucleic acid sequence.  
KM Mouse; murine; tumour; cancer; metastatic sequence; detection;  
KM diagnosis; treatment; metastasis; hyperplasia; dysplasia;  
KM hypertrophy; screening; ss.  
OS Mus musculus.  
PN WO9718454-A2.  
PD 22-MAY-1997.  
PF 15-NOV-1996: U18567.  
PR 30-JAN-1996: US-594031.  
PR 16-NOV-1995: US-006838.  
PA (THOM/) THOMPSON T.  
PI Thompson T.  
DR WPI: 97-289397/26.  
PT Identifying tumour metastatic sequences - by introducing transfected  
PT cells into host mammal and analysing primary and metastatic  
PT sequences by differential display PCR  
PS Disclosure: Fig 12B; 102pp; English.  
CC Mouse Orogenital Sinus (UGS) tissue was isolated from 17 day old  
CC mouse embryos. The UGS cells were infected with retroviruses,  
CC cultured and implanted under the renal capsule of mice.  
CC Reconstructions were harvested 5 weeks later. When they showed  
CC signs of distress from the tumour burden. Metastasised tumours were  
CC isolated from a site outside the renal capsule. RNA was isolated  
CC from primary tumours and metastases, reverse transcribed and  
CC subjected to differential display PCR. The sequences were analysed  
CC to obtain metastatic sequences, e.g. the present sequence. The  
CC method can be used to detect, diagnose and treat disorders related  
CC to metastasis, or treat malignant or non-malignant disorders, e.g.  
CC hyperplasia, dysplasia and hypertrophy. The metastatic sequence can  
CC be used to screen a biological sample for metastasis, and it or its  
CC expression product may also be used to treat a metastatic disorder.  
SQ Sequence 371 BP; 103 A; 75 C; 84 G; 109 T;

Query Match 9.7%; Score 28; DB 1; Length 371;  
Best Local Similarity 55.0%; Pred. No. 3.5;  
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 155 AAGATGACGAGAGCTTACAGACATATGCTTCAGATCAATGAGACATGAGCCTCAT 214  
DB 75 AAGATGCTAGATGTTTACACAGCTGTGAACAGGGCTCAATGATCGTGATGCTTG 134  
OY 215 GATGCTTTGGGCCCTTCATCAGAGAAATGAGAGCTGCT 254  
DB 135 AGAATGCTTTCTTGTGTGACATCAGAGAACTGGAATGTT 174

RESULT 6  
V80591  
ID V80591 standard; cDNA; 3531 BP.  
AC V80591;  
DE 01-MAR-1999 (first entry)  
DE Kidney injury associated molecule HM036 cDNA clone.  
KM Kidney injury associated molecule; kidney injury related molecule;  
KM KIM; tissue growth promotion; regeneration; renal condition;  
KM acute renal failure; acute nephritis; tumour; ds.  
OS Rattus sp.  
FT Key location/Qualifiers  
FT misc\_feature complement (2541..2575)  
FT /tag= a  
FT /label= SAC\_24406  
PN WO9853071-A1.  
PD 26-NOV-1998.  
PF 22-MAY-1998: U10547.  
PR 23-MAY-1997: US-047491.  
PR 23-MAY-1997: US-047490.  
PA (BIOJ) BIOGEN INC.  
PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;  
DR WPI: 99-045312/04.  
PT Kidney injury-associated molecule, KIM, polypeptides - upregulated  
PT in injured or regenerating tissues, useful to promote tissue growth

PT and regeneration, especially to treat renal conditions  
PS Claim 9; Page 70-71; 213pp; English.  
CC The present sequence represents a kidney injury associated molecule  
CC (KIM) cDNA clone. KIM proteins can be administered therapeutically  
CC by expressing KIM encoding polynucleotides, to promote growth and/or  
CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins  
CC are upregulated in injured or regenerating (especially renal) tissues.  
CC KIM fusion proteins, conjugates, antibodies and vectors can also be used  
CC therapeutically, e.g. these or the KIM proteins may be included with an  
CC acceptable carrier in pharmaceutical compositions, useful for therapy/  
CC prophylaxis of conditions associated with dysfunction/dysregulation of  
CC KIM genes or proteins, especially renal diseases or impairments of renal  
CC function in humans (e.g. acute renal failure, acute nephritis). The  
CC polynucleotides can be used to produce antisense sequences which, when  
CC internalised into cells, can disrupt expression of a cellular KIM gene,  
CC also useful in therapy (e.g. to block the growth of tumours dependent on  
CC KIM for growth) or compositions. The proteins and polynucleotides are  
CC useful diagnostically e.g. to detect and quantify renal injury/disease  
CC (indicative of increased risk, or presence of, renal injury or impaired  
CC function), or abnormal responses to tissue injury (indicative of  
CC increased risk, or presence of, an autoimmune response or abnormal  
CC tissue growth arising from/affecting renal tissue). The proteins can  
CC also be used to locate KIM-producing cells (especially specific loci,  
CC e.g. tissue masses abnormally producing/expressing KIM such as tumours  
CC arising from/affecting renal tissue), by contacting cells with an  
CC imageable KIM-binding reagent and imaging reagent accumulation.  
SQ Sequence 3531 BP; 846 A; 795 C; 732 G; 1158 T;

Query Match 9.7%; Score 28; DB 1; Length 3531;  
Best Local Similarity 55.0%; Pred. No. 9.6;  
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 113 GCTTGGGCCCTGCTGACCTCAGACAGAGAGAGAGCTCCAAATGACAGAGCTTAA 172  
DB 2738 GCCGTGGGTAAGCTGATGTTTCCACCTCGAGAGCTCTACATGCTGTGACATGTATA 2797  
OY 173 CAGACATATGCTTCAGATCAAGTTCAGAAATGGCCTTA 212  
DB 2798 TACAATATCCCTCGGATCCTGTCATGATACAGCAGCA 2837

RESULT 7  
Q59984  
ID Q59984 standard; cDNA; 302 BP.  
AC Q59984;  
DE 16-MAR-1994 (first entry)  
DE Human brain Expressed Sequence Tag EST01961.  
KM Gene transcription product; genetic markers; tagging; in vivo;  
KM transcription; mapping; locations; chromosomes; chromosomal; ss.  
OS Homo sapiens.  
PN WO9316178-A.  
PD 19-AUG-1993.  
PF 12-FEB-1993: U01294.  
PR 12-FEB-1992: US-837195.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
PI Adams MD, Moreno RF, Venter CJ;  
DR WPI: 93-272882/34.  
PT Enriched oligonucleotides and corresp. sequences - used as  
PT markers for human genes transcribed in-vivo, facilitate tagging  
PT of most human genes  
PS Example 4; Page 262; 500pp; English.  
CC The Expressed Sequence Tag was isolated from a human brain cDNA  
CC library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes  
CC on chromosomes, for individual or forensic identification, for mapping  
CC locations of disease-associated genes, for identification of tissue  
CC type, and for prep. of antisense sequences, probes and constructs.  
CC EST01961 has been putatively identified as encoding a Notch/Notch  
CC protein. See also 059041-061440  
SQ Sequence 302 BP; 72 A; 79 C; 69 G; 77 T;

OS Tomato spotted wilt virus.  
 PN W09629420-A1.  
 PD 26-SEP-1996.  
 PF 22-MAR-1996; E01271.  
 PR 23-MAR-1995; GB-005907.  
 PA (SANO ) SANDOZ LTD.  
 PA (SANO ) SANDOZ PATENT GMBH.  
 PA (SANO ) SANDOZ-ERFINDUNGEN VERW GMBH.  
 PI De Haan PT;  
 DR WPI: 96-443194/44.  
 PT Recombinant nucleic acid contg. modified tospovirus sequence - used to prepare plants resistant or tolerant to tospovirus  
 PS Claim 1; Page 14-16; 38pp; English.  
 CC A DNA sequence (T38305) is complementary to one which encodes the tomato spotted wilt virus M-RNA sequence which itself encodes the putative viral movement protein. A recombinant nucleic acid contains a transcriptional regulatory region and, under its control, the tospovirus sequence, pref. modified by removal or inactivation of the start codon, and insertion or deletion of the nucleotides 3' of the start codon to alter the reading frame of the encoded RNA. A plant having such a construct stably integrated and expressed in its genome is resistant to, or tolerant of, tospoviruses.  
 CC Sequence 4821 BP: 1569 A: 850 C: 870 G: 1532 T;  
 SQ

Query Match 10.0%; Score 29; DB 1; Length 4821;  
 Best Local Similarity 63.8%; Pred. No. 5.1;  
 Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 180 ATGCTTCAGATCAGAGTTCAGAACATGCGCTTCATGATGCTTTGGGCGCTTCATGAG 239  
 DB 476 ATGTTTATATACAGAGTGAACAAATGCTGATATATCTGCTTGTGATTCAGAG 535  
 QY 240 GAATGAG 248  
 DB 536 AAGGTTAG 544

RESULT 3  
 V89065  
 ID V89065 standard; cDNA; 414 BP.  
 AC V89065.  
 DT 15-FEB-1999 (first entry)  
 DE EST clone B253.  
 KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
 KW gene therapy; ss.  
 OS Homo sapiens.  
 PN W09845436-A2.  
 PD 15-OCT-1998.  
 PF 10-APR-1998; U06955.  
 PR 10-APR-1997; US-838821.  
 PA (CEMY ) GENETICS INST INC.  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Meisberg D,  
 PI Racie LA, Spaulding V, Treacy M;  
 PT WPI: 99-070077/06.  
 PT New polynucleotides encoding human secreted proteins - derived from e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries.  
 PS Claim 1; Page 98; 618pp; English.  
 CC The present sequence represents a human expressed sequence tag (EST).  
 CC The polynucleotide, which is a secreted EST, and the encoded protein are predicted to have useful biological activities which would make conditions suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional activity, immune stimulating or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, tumour inhibition

CC activity. The polynucleotide may also be useful for gene therapy.  
 SQ Sequence 414 BP: 98 A: 90 C: 100 G: 126 T;

Query Match 9.9%; Score 28.6; DB 1; Length 414;  
 Best Local Similarity 55.6%; Pred. No. 2.3;  
 Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 130 AGCTCAGAACTAGAGAGCTGTCCAAAATGACAGAGGTTTACAGACATAGCTTCAGA 189  
 DB 224 AGGTTTGAATCCAGACATATGCTGAACAGCGGAGCTTCACTGATTCACACAGA 283  
 QY 190 TCAAGTTCAGAACATGCGCTTCATGATGCTTTGGGCGC 228  
 DB 284 GAATTTATAGCAGAGTGCGTGCAGCATGTGCTTGAGGCC 322

RESULT 4  
 005733/c  
 ID 005733 standard; DNA; 753 BP.  
 AC 005733;  
 DT 07-JAN-1991 (first entry)  
 DE Glutathione-S-transferase (GST)-7; liver fluke infection vaccine;  
 KW glutathione-S-transferase (GST)-7; liver fluke infection vaccine;  
 KW ss.  
 OS Fasciola hepatica.  
 FH Key location/Qualifiers  
 FT poly-a-site 745..745  
 FT /\*tag- a  
 PN W09008819-A.  
 PD 09-AUG-1990.  
 PF 31-JAN-1990; A00027.  
 PR 31-JAN-1989; AD-002481.  
 PA (DARA-) DARATECH PTY LTD.  
 PI Crameri SL, Panaccio M, Wilson R, Wijffels GL, Spithill TW,  
 PI Campbell NJ, Thompson C, Sexton JL;  
 DR WPI: 90-260928/34.  
 DR P-PSDB; R06499.  
 PT Vaccine against liver fluke infection in ruminants, esp. sheep -  
 PT comprises glutathione-S-transferase extracted from adult fasciola hepatica worms.  
 PS Claim 9; Fig 15; 44pp; English.  
 CC A cDNA library was constructed from total RNA of adult worms of the Compton strain of Fasciola hepatica in either lambda gtl1 or lambda ZAP phage vectors. The library was screened with rabbit antiserum raised to purified GSTs of F.hepatica. Clone GST-7 was isolated from a lambda ZAP library and was used to isolate other homologous sequences. It contains a poly(A) tail indicating the 3' end of mRNA has been cloned. The protein derived from this sequence begins 7 amino acids from the N-terminus of GST(Fh2a) and is homologous to it.  
 CC This sequence can be used in a vaccine for preventative treatment of liver fluke in ruminants.  
 CC See also 005732, 005734-005736 and R06503-R06506.  
 SQ Sequence 753 BP: 219 A: 146 C: 176 G: 212 T;

Query Match 9.8%; Score 28.2; DB 1; Length 753;  
 Best Local Similarity 55.7%; Pred. No. 4.1;  
 Matches 54; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 166 AGGTTTCAGACATATGCTTCAGATCAAGTTTCAGAACAAATGGCTCTTATGCTTTGG 225  
 DB 675 AGATTGACAGAAATCCATCTGTGATTCAGCTGAACAAAGATCTCTCAAGCAGGTG 616  
 QY 226 GGCCCTTCATTCAGAAATGAGAGCTGCTCCAGCGC 262  
 DB 615 AGCGTTCACACCCAAAGAGAGATCCAGAGCTGC 579

RESULT 5  
 T69307  
 ID T69307 standard; cDNA; 371 BP.

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:02:41 ; Search time 921.96 Seconds  
(without alignments)  
78.426 Million cell updates/sec

Title: US-09-049-696-7

Perfect score: 289  
Sequence: 1 GAAATATCCAACTGATGGAT .....AGCTTGAGATGAAGGATTA 289

Scoring table: IDENTITY\_NUC

Searched: 311585 segs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31.4	10.9	678	1	EST clone BK296. N
2	29	10.0	4821	1	TOSPOVIRUS M-RNA S
3	28.6	9.9	414	1	EST clone B253. Ne
4	28.2	9.8	753	1	Glutathione-S-tran
5	28	9.7	371	1	Murine metastatic
6	28	9.7	3531	1	Kidney injury asso
7	27.8	9.6	302	1	Human Brain Expres
8	27.8	9.6	110000	1	Continuation (4 of
9	27.8	9.6	110000	1	Continuation (5 of
10	27.2	9.5	8726	1	Enterococcus faeca
11	27.2	9.4	4062	1	C3G protein gene.
12	27.2	9.4	631	1	S. pneumoniae dete
13	27.2	9.4	352	1	Streptococcus pneu
14	27.2	9.4	10254	1	Streptococcus pneu
15	27.2	9.4	1660	1	Streptococcus pneu
16	27	9.3	7462	1	Jaagsiekte retrovi
17	27	9.3	4992	1	Cancer associated
18	27	9.3	1623	1	Staphylococcus aur
19	26.8	9.3	8133	1	C. difficile toxin
20	26.8	9.3	8133	1	Clostridium diffic
21	26.8	9.3	5360	1	Enterococcus faeca
22	26.8	9.3	644	1	Enterococcus faeca
23	26.8	9.3	648	1	Enterococcus faeca
24	26.8	9.3	394	1	Human secreted pro
25	26.6	9.2	2305	1	Annexin XI gene in
26	26.6	9.2	2311	1	Annexin XI gene in
27	26.6	9.2	3872	1	Sockeye salmon gro
28	26.6	9.2	371	1	Heat shock inducib
29	26.6	9.2	371	1	Heat shock inducib
30	26.6	9.2	371	1	Heat shock inducib
31	26.6	9.2	371	1	Heat shock inducib
32	26.6	9.2	4232	1	DNA molecule encod
33	26.6	9.2	3872	1	Enterococcus faeca
34	26.6	9.2	1958	1	Oncofyrnchus nerka
35	26.6	9.2	1958	1	Lung cancer associ
36	26.4	9.1	2706	1	L-lactic acid dehyd
37	26.4	9.1	882	1	Rat oploph recept
38	26.4	9.1	110000	1	Type A2 ACCase gen
39	26.4	9.1	3495	1	Continuation (2 of
40	26.4	9.1	1220	1	GlutRNAGLN amidotr
41	26.4	9.1	5549	1	Human uncoupling p
42	26.4	9.1	13340	1	Staphylococcus aur
43	26.4	9.1	110000	1	O. longistamnatu
					Continuation (2 of

44 26.2 9.1 1581 1 N82012 Catalase gene. Hum  
45 26.2 9.1 636 1 V29365 Calcium ion channel

## ALIGNMENTS

RESULT 1  
V86928  
ID V86928 standard; cDNA; 678 BP.  
AC V86928;  
DT 27-APR-1999 (first entry)  
DE EST clone BK296.  
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
KW tissue growth; actinin; inhibin; tumour invasion suppressor; EST; human;  
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
OS Homo sapiens.  
PN W09845435-A2.  
PD 15-OCT-1998.  
PF 10-APR-1998; U06954.  
PR 10-APR-1997; US-835913.  
PA (GENE) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racle DA, Spaulding V, Treacy M;  
DR WPI; 99-070076/06.  
PT New polynucleotides encoding human secreted proteins - derived from  
PT ovary, pituitary, kidney, foetal lung, placenta, testes, brain,  
PT Claim 1, Page 394-395, 633pp, English.  
PS This sequence represents an expressed sequence tag (EST), and is a  
CC polynucleotide of the invention. The polynucleotides of the invention are  
CC all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC actinin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The EST sequences are also stated to be useful for gene  
CC therapy.  
SQ Sequence 678 BP; 120 A; 172 C; 198 G; 188 T;

Query Match 10.9%; Score 31.4; DB 1; Length 678;  
Best Local Similarity 52.7%; Pred. No. 0.33;  
Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 146 GAGCTGTCACAAATGACAGAGGTTTACAGATATGCTTCAGATCAAGTTCAGAACAT 205  
DB 263 GATTGTGTCACAAATGAGGTTTACAGATATGCTTCAGATCAAGTTCAGAACAT 322  
QY 206 GGCCTCATGATGATCTTTGGGCCCCCTTCATCAGGAAATGAGCTGTCTCAGGCGCTC 265  
DB 323 GGGATGACAGGCTCTGAGAGGAGCCCTCTTCAGAGACAGGCGCCCTGAGGAGTCC 382  
QY 266 ATCCAGCTT 274  
DB 383 ACTCAGCTT 391  
RESULT 2  
T8305  
ID T8305 standard; DNA; 4821 BP.  
AC T8305;  
DT 03-JAN-1997 (first entry)  
DE TOSPOVIRUS M-RNA sequence.  
KW TOSPOVIRUS; disease resistance; crop protection; transgenic plant;  
KW tomato chlorotic spot virus; groundnut ringspot virus;  
KW Impatiens necrotic spot virus; viral movement protein; ds.



PF 22-JUN-1990; U03586.  
 PR 23-JUN-1989; US-370673.  
 PA (GETH ) GENENTECH INC.  
 PI Chang M, Goeddel D, Lowe D;  
 DR WPI: 91-036711/05.  
 DR P-PSDB: R10399.  
 PT Natriuretic protein receptor B - for diagnosis and treatment of  
 PT Kidney failure, heart failure, hyperaldosteronism, glaucoma etc.  
 PS Claim 3; Fig 1; 49p; English.  
 CC The sequence encodes natriuretic peptide receptor B, NPRB, having  
 CC guanyl cyclase (GC) activity and protein kinase activity. The DNA  
 CC can be inserted into expression vectors for the prodn. of the  
 CC protein, opt. after being mutated to produce NPRB analogues. The  
 CC sequence can also be used to screen for mRNA encoding NPRB for  
 CC diagnostic purposes. The protein itself (or variants) can be used  
 CC in treatment of natriuretic peptide disorders, and also to isolate  
 CC peptides using affinity chromatography. Antibodies with affinity  
 CC for NPRB can also be prepd.  
 SQ Sequence 3697 BP; 793 A; 1048 C; 981 G; 875 T;

Query Match 5.2%; Score 15; DB 1; Length 3697;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 121 GCCCTCTGCAGCTCA 135  
 ||||||||||||  
 Db 3576 GCCCTCTGCAGCTCA 3590

Search completed: August 5, 1999, 17:40:25  
 Job time: 6260 sec



```
FT      /*tag- 2
FT      /label= MG070
FT      /note= "Previously identified as MORF-20136, the
FT      encoded protein shows 34.8 percentage

Query Match
Best Local Similarity 100.0%; Score 16; DB 1; Length 110000;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      47 GAAGACACACTATA 62
DB      30560 GAAGACACACTATA 30575

RESULT 10
V49253/c
ID      V49253 standard: DNA: 2218 BP.
AC      V49253:
DE      28-OCT-1998 (first entry)
KW      Mouse delta opiate receptor gene.
KW      Mouse; delta opiate receptor; transgenic animal; mammal; identification;
KW      exon; nervous tissue; pain; drug addiction; transplant rejection;
KW      immunosuppressant; analgesic; morphine; side effect; ds.
OS      Mus sp.
FH      Location/Qualifiers
FT      key
FT      CDS
FT      58..1176
FT      /tag= a
FT      /product= "delta opiate receptor"

MO9802534-A2.
PD      22-JAN-1998
PR      11-JUL-1997; F01282
PR      15-JUL-1996; FR-008810.
PA      (CNRS ) CENT NAT RECH SCI.
PI      Dierich A, Kleiffer BL, Lemeur M, Matthes HMD, Simonin FH;
DR      WPI: 98-110582/10.
DR      P-PSDB: W44938.
PT      Transgenic animals defective in one type of opiod receptor - used
PT      to identify agents for treatment of pain, drug addiction and
PT      transplant rejection, lacking side effects of known opiate(s)
PS      Disclosure: Fig 12; 58pp; French.
CC      This sequence represents the gene encoding the mouse delta opiate
CC      receptor protein. The sequence is used to generate a transgenic
CC      non-human mammal for identifying agents for treating disorders
CC      associated with opiate receptors. In the mammal, the expression of
CC      the gene encoding the opiate receptor is modified, particularly by
CC      the deletion of an exon and/or insertion of a marker gene, e.g. the
CC      neomycin resistance gene, into the sequence. Especially the expression
CC      of the gene is altered in nervous tissue. The agents are potentially
CC      useful for treating severe pain (chronic or acute), drug addiction and/or
CC      prevention or treatment of transplant rejection (as immunosuppressants).
CC      The method may isolate and identify powerful analgesics that lack
CC      morphine-like side effects.
SQ      Sequence 2218 BP; 460 A; 648 C; 650 G; 460 T;

Query Match
Best Local Similarity 100.0%; Score 16; DB 1; Length 2218;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      217 TGCTTTGGGCGCCTT 232
DB      2133 TGCTTTGGGCGCCTT 2118

RESULT 11
V70360/c
ID      V70360 standard: cDNA: 2981 BP.
AC      V70360:
DE      08-FEB-1999 (first entry)
KW      Human ehb21 encoding cDNA.
KW      Human; cytoplasmic protein; EH-containing protein; eps15; eps15R;
KW      eps15*homology; intracellular interaction; EH domain binding specificity;
KW      signal transducer; NPF motif; h-NUMB; h-NUMB-R; h-RAB; h-RAB-R; ehb3;
```

```
KW      ehb10; ehb21; cell proliferation; diagnosis; detection; ss.
OS      Homo sapiens.
FH      Location/Qualifiers
FT      key
FT      CDS
FT      88..456
FT      /*tag= a

MO9846744-A1.
PD      22-OCT-1998
PR      06-APR-1998; IT0077.
PR      15-APR-1997; IT-MI0868.
PA      (EUNON-) 1ST EURO DI ONCOLOGIA SRL.
PI      Di Fiore PP, Doria M, Pellicci PG, Salcini AE;
DR      WPI: 98-584574/50.
DR      P-PSDB: W83219.
PT      New isolated EH domain binding proteins and peptide(s) - obtained
PT      using signal transducers eps15 and eps15R containing EH domains by
PT      detecting specific binding activity.
PS      Claim 4; Page 54-56; 90pp; English.
CC      The present invention describes a new intracellular interactor and novel
CC      protein:protein interaction (EH) domain binding protein having (parts of)
CC      one of the following human derived sequences designated: (i) h-NUMB;
CC      (ii) h-NUMB-R; (iii) h-RAB-R; (iv) ehb3; (v) ehb10; or (vi) ehb21. The
CC      present invention also describes peptides containing at least one NPF
CC      (Asp-Pro-Phe) motif, able to bind to a protein with at least one EH
CC      domain. The proteins with a NPF-containing peptide or the peptides
CC      themselves can be used to identify and purify EH containing proteins.
CC      Antisense RNA, complementary to mRNA encoding h-NUMB, h-NUMB-R, h-RAB-R,
CC      ehb3, ehb10, or ehb21, can be used for diagnostic and therapeutic uses.
CC      Products from the present invention can also be used to develop agents
CC      for use in control of cell proliferation. The present sequence encodes
CC      human ehb21.
SQ      Sequence 2981 BP; 705 A; 780 C; 722 G; 774 T;

Query Match
Best Local Similarity 100.0%; Score 16; DB 1; Length 2981;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      227 GCCCTTCATCAGGAA 242
DB      593 GCCCTTCATCAGGAA 578

RESULT 12
X23519/c
ID      X23519 standard: DNA: 44453 BP.
AC      X23519:
DE      23-JUN-1999 (first entry)
KW      Human kidney aminopeptidase P genomic DNA fragment 3.
KW      Aminopeptidase; human; Amp; gene therapy; treatment; Amp-deficiency;
KW      prenatal diagnosis; angioedema; antihypertensive agent; atherosclerosis;
KW      arterial stenosis; industrial protein feed; malabsorption syndrome;
KW      proteolaceous waste degradation; additive; immunohistochemistry; ss.
OS      Homo sapiens.
PD      MO9911799-A2.
PD      11-MAR-1999
PR      02-SEP-1998; U18426.
PR      02-SEP-1997; US-057854.
PA      (MEDICAL COLLEGE GEORGIA RES INST.
PI      Ryan JW, Sprinkle TTC, Venema RC;
DR      WPI: 99-205193/17.
PT      Nucleic acid encoding human aminopeptidase P
PS      Claim 13; Page 139-165; 201pp; English.
CC      This invention describes the isolation of a novel human aminopeptidase P
CC      (Amp). This protein is used to produce recombinant Amp and can be used
CC      for gene therapy for treating Amp-deficiency conditions. Its fragments
CC      are used as primers and probes to identify patients with homozygous and
CC      heterozygous Amp deficiency, including prenatal diagnosis (patients
CC      defective in Amp are at risk of developing angioedema if treated with
CC      angiotensin-converting enzyme inhibitors), also as antisense inhibitors
CC      in cases of excessive Amp expression. The product of the invention is
CC      also used to identify Amp-expressing sequences in other animals and to
CC      generate transgenic animals, and comparisons of genomic sequences are
CC      used to detect mutations. Amp inhibitors are potentially useful as
```

FT and MORF-20081, the encoded protein shows 33.04  
FT percentage identity to 5,10-methylene-tetra-  
FT hydrofolate dehydrogenase (fold) from *E. coli*"  
FT 17474..19243  
FT /tag- f  
FT /label- MG015  
FT /note- "Previously identified as MORF-20084, the  
FT encoded protein shows 32.23 percentage  
FT identity to transport ATP-binding protein  
FT (msbA) from *E. coli*"  
FT 26478..27344  
FT /tag- g  
FT /label- MG023  
FT /note- "Previously identified as MORF-20092, the  
FT encoded protein shows 45.96 percentage  
FT identity to fructose-bisphosphate aldolase  
FT (fbs) from *B. subtilis*"  
FT 27345..28448  
FT /tag- h  
FT /label- MG024  
FT /note- "Previously identified as MORF-19826 and  
FT MORF-20093, the encoded protein shows 46.84  
FT percentage identity to GTP-binding protein  
FT from *E. coli*"  
FT 36987..38978  
FT /tag- i  
FT /label- MG032  
FT /note- "Previously identified as MORF-20099, the  
FT encoded protein shows 26.82 percentage  
FT identity to ATP-dependent nuclease (adda)  
FT from *B. subtilis*"  
FT 39242..39904  
FT /tag- j  
FT /label- MG033  
FT /note- "Previously identified as MORF-20100, the  
FT encoded protein shows 35.90 percentage  
FT identity to glycerol uptake facilitator  
FT (glpf) from *B. subtilis*"  
FT complement (39873..40514)  
FT /tag- k  
FT /label- MG034  
FT /note- "Previously identified as MORF-20101, the  
FT encoded protein shows 48.13 percentage  
FT identity to thymidylate kinase (tdk)  
FT from *B. subtilis*"  
FT 40543..41787  
FT /tag- l  
FT /label- MG035  
FT /note- "Previously identified as MORF-20102, the  
FT encoded protein shows 30.71 percentage  
FT identity to histidyl-tRNA synthetase (hiss)  
FT from *Mycobacterium leprae*"  
FT complement (44751..46277)  
FT /tag- m  
FT /label- MG038  
FT /note- "Previously identified as MORF-20105, the  
FT encoded protein shows 46.83 percentage  
FT identity to glycerol kinase (glpk)  
FT from *E. coli*"  
FT complement (46268..47422)  
FT /tag- n  
FT /label- MG039  
FT /note- "Previously identified as MORF-19831 and  
FT MORF-20106, the encoded protein shows 43.20  
FT percentage identity to glycerol-3-phosphate  
FT dehydrogenase (GUT2) from *S. cerevisiae*"  
FT 49377..49643  
FT /tag- o  
FT /label- MG041  
FT /note- "The encoded protein shows 48.86 percentage  
FT identity to phosphohistidinophosphatase-  
FT phosphotransferase (ptsh) from *Mycoplasma*  
FT *capricolum*"

FT cds  
FT 50060..51520  
FT /tag- p  
FT /label- MG042  
FT /note- "Previously identified as MORF-19832 and  
FT MORF-20108, the encoded protein shows 41.92  
FT percentage identity to spermidine/  
FT putrescine transport ATP-binding protein  
FT (potA) from *E. coli*"  
FT 51525..52382  
FT /tag- q  
FT /label- MG043  
FT /note- "Previously identified as MORF-20110, the  
FT encoded protein shows 26.51 percentage  
FT identity to spermidine/putrescine transport  
FT system permease protein (potB) from *E. coli*"  
FT 52366..53220  
FT /tag- r  
FT /label- MG044  
FT /note- "Previously identified as MORF-20111, the  
FT encoded protein shows 29.45 percentage  
FT identity to spermidine/putrescine transport  
FT system permease protein C (potC) from *E. coli*"  
FT 54658..55605  
FT /tag- s  
FT /label- MG046  
FT /note- "Previously identified as MORF-20112, the  
FT encoded protein shows 36.60 percentage  
FT identity to salicoglycoprotease (gsp)  
FT from *Pasteurella haemolytica*"  
FT complement (56970..58310)  
FT /tag- t  
FT /label- MG048  
FT /note- "Previously identified as MORF-19834,  
FT MORF-20114 and MORF-20115, the encoded protein  
FT shows 43.02 percentage identity to signal  
FT recognition particle protein (fth) from *B.*  
FT *subtilis*"  
FT 58117..59079  
FT /tag- u  
FT /label- MG049  
FT /note- "Previously identified as MORF-20114 and  
FT MORF-20115, the encoded protein shows 44.78  
FT percentage identity to putine-nucleoside  
FT phosphorylase (deop) from *E. coli*"  
FT 59083..59754  
FT /tag- v  
FT /label- MG050  
FT /note- "Previously identified as MORF-20117, the  
FT encoded protein shows 83.03 percentage  
FT identity to deoxyribose-phosphate aldolase  
FT (deoc) from *Mycoplasma pneumoniae*"  
FT complement (64898..65731)  
FT /tag- w  
FT /label- MG056  
FT /note- "Previously identified as MORF-20122, the  
FT encoded protein shows 30.25 percent  
FT identity to the protein disclosed in  
FT GB:D26185..99 from *B. subtilis*"  
FT complement (65713..66249)  
FT /tag- x  
FT /label- MG057  
FT /note- "Previously identified as MORF-20123, the  
FT encoded protein shows 38.90 percentage  
FT identity to the protein disclosed in  
FT GB:D26185..104 from *B. subtilis*"  
FT 81047..82597  
FT /tag- y  
FT /label- MG067  
FT /note- "Previously identified as MORF-19845, the  
FT encoded protein shows 28.84 percentage  
FT identity to glutamic acid specific protease  
FT (Spase) from *Staphylococcus aureus*"  
FT 91065..91919  
FT cds

```

FT cds 12.1130
FT /tag- a
FT /product- mouse delta opioid receptor
PN MO9428132-A.
PD 08-DEC-1994.
PE 20-MAY-1994. U05747.
PR 20-MAY-1993. US-066296.
PR 30-JUL-1993. US-100694.
PR 05-NOV-1993. US-147592.
PA (ARCH-) ARCH DEV CORP.
PI Bell GI, Reisine T, Yasuda K:
DR WPI: 95-022804/03.
DR P-PSDB: R67670.
PT Polynucleotides and peptides derived from opioid receptor
PT polypeptides - for use in therapeutic compositions and in
PT screening assays for useful drug substances.
PS Claim 6: Page 215-221. 300pp. English.
CC The nucleotide sequence of the novel mouse delta opioid receptor gene
CC MORL1. The gene was isolated from a mouse brain cDNA library using a
CC fragment (amplified from the cDNA library with primers Q75929-30) as a
CC probe. The primers are based on the conserved sequences present in the
CC second and third transmembrane domains of somatostatin (SRIF) receptor
CC subtypes SSTR1, SSTR2 and SSTR3. The 1.3 kb EcoRI-SacI fragment from the
CC mouse delta opioid receptor clone, lambda msl-2, was subcloned into the
CC CMV promoter-based expression vector pCMV-6c. The resultant construct
CC pCMV-msl-2 was transfected into COS-1 cells for protein production. The
CC gene encoding the opioid receptor can be used to produce complete,
CC truncated or chimeric opioid receptor proteins. The opioid receptors
CC thus produced are useful for the development of novel assays designed to
CC select or improve substances, capable of interacting with the opioid
CC receptor proteins, for use in diagnosis, drug design and therapeutic
CC applications.
SQ Sequence 2272 BP: 485 A; 665 C; 650 G; 472 T;

Query Match 5.5%; Score 16; DB 1; Length 2272;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 217 TGCTTTGGGGCCCTT 232
DB 2073 TGCTTTGGGGCCCTT 2058

RESULT 8
T21900/c
ID T21900 standard; cDNA to mRNA; 273 BP.
AC T21900;
DT 14-AUG-1996 (first entry)
DE Human gene signature HUMGS03442.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PE 11-NOV-1994. J01916.
PR 12-NOV-1993. JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K:
PI Matsubara K, Okubo K:
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues.
PS Claim 1: Page 996. 2245pp. Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in 119001-126837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the

```

```

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 273 BP: 58 A; 74 C; 71 G; 70 T;

Query Match 5.5%; Score 16; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 GGAAGCAACACTATA 61
DB 34 GGAAGCAACACTATA 19

RESULT 9
T58840.0
WP Sequence split into 6 fragments LOCUS T58840 Accession T58840
WP Fragment Name Begin End
WP T58840_0 1 110000
WP T58840_1 100001 210000
WP T58840_2 200001 310000
WP T58840_3 300001 410000
WP T58840_4 400001 510000
WP T58840_5 500001 580073
ID T58840 standard; DNA; 580073 BP.
AC T58840;
DT 27-MAR-1997 (first entry)
DE Mycoplasma genitalium genome.
KW M. genitalium; DNAA; DNA gyrase; origin of replication;
KW megabase shotgun sequencing method; open reading frame; ORF; ss.
OS Mycoplasma genitalium.
FH Key Location/Qualifiers
FT cds 8552..9184
FT /tag- a
FT /label- MG006
FT /note- "Previously identified as MORF-20076, the
FT encoded protein shows 27.59 percentage
FT identity to thymidylate kinase (CDC8)
FT from Saccharomyces cerevisiae"
FT 11252..12040
FT /tag- b
FT /label- MG009
FT /note- "Previously identified as MORF-20078, the
FT encoded protein shows 35.43 percentage
FT identity to the Bacillus subtilis hypothetical
FT protein covered in accession number
FT GB:D26185.102"
FT 12069..12725
FT /tag- c
FT /label- MG010
FT /note- "Previously identified as MORF-20079, the
FT encoded protein shows 25.73 percentage
FT identity to DNA primase (dnaE) from
FT Clostridium acetobutylicum"
FT complement (13570..14247)
FT /tag- d
FT /label- MG012
FT /note- "Previously identified as MORF-20080, the
FT encoded protein shows 31.50 percentage
FT identity to the ribosomal protein S6
FT modification protein (rimK) from Escherichia
FT coli"
FT complement (14396..15217)
FT /tag- e
FT /label- MG013
FT /note- "Previously identified as MORF-19823, MORF-20080

```

FT	variation	4071	/tag= i
FT		/label=	Polymorphic_site
FT	variation	4383	/tag= j
FT		/label=	Polymorphic_site
FT	variation	4462	/tag= k
FT		/label=	Polymorphic_site
FT	variation	4494	/tag= l
FT		/label=	Polymorphic_site
FT	variation	6867	/tag= m
FT		/label=	Polymorphic_site
FT	variation	7692	/tag= n
FT		/label=	Polymorphic_site
FT	variation	8940	/tag= o
FT		/label=	Polymorphic_site
FT	variation	9192	/tag= p
FT		/label=	Polymorphic_site
FT	variation	9585	/tag= q
FT		/label=	Polymorphic_site
FT	variation	9600	/tag= r
FT		/label=	Polymorphic_site
FT	variation	9951	/tag= s
FT		/label=	Polymorphic_site
FT	variation	10111	/tag= t
FT		/label=	Polymorphic_site
FT	variation	11250	/tag= u
FT		/label=	Polymorphic_site
FT	variation	12300	/tag= v
FT		/label=	Polymorphic_site
FT	variation	14007	/tag= w
FT		/label=	Polymorphic_site
FT	variation	15355	/tag= x
FT		/label=	Polymorphic_site
PN	Mo9211387-A.		
PD	09-JUL-1992.		
PF	20-DEC-1991.	CA0457.	
PR	21-DEC-1990.	GB-027869.	
PR	20-MAY-1991.	GB-010865.	
PR	09-SEP-1991.	GB-019250.	
PA	(UYGU-) UNIV GUERLPH.		
PA	(UTOR ) UNIV TORONTO INNOVATIONS FOUND.		
PI	MacLennan DH, O'Brien PJ;		
DR	P-PSDB; R25450.		
DR	WP1; 92-250106/30.		
PT	Purified DNA mol. for diagnosis of porcine malignant hyperthermia		
PT	- compliss DNA sequence encoding normal or mutant ryanodine		
PT	receptor with specified endonuclease restriction map		
PS	Disclosure; Fig 2: 96pp; English.		
CC	The sequence given is the mutant pig ryanodine receptor (RyR1) gene		
CC	from swine cDNA. The polymorphic sites were observed in comparisons		
CC	of Pietrain and Yorkshire breeds. There are 17 polymorphisms between		
CC	the two breeds. The polymorphism at position 1972 causes a mutation		
CC	from Arg to Cys and this is thought to be the molecular basis of		
CC	porcine malignant hyperthermia (MH). This mutation lies within the		
CC	region of RyR1 that is concerned with the binding of regulators of Ca2+		
CC	release channel gating. Analysis of surrounding sequences suggests		
CC	that this mutation lies within a beta strand domain comprising roughly		
CC	of amino acids 520 to 830. RyR1 is the calcium release channel of the		
CC	sarcoplasmic reticulum and is a large protein which spans the gap		

CC between the transverse tubule and the sarcoplasmic reticulum. The  
CC channel is activated by ATP, calcium, caffeine, and micro-molar  
ryanodine. It is inhibited by ruthenium red, tetracaine, calmodulin,  
CC high Mg<sup>2+</sup> and ryanodine.  
SQ Sequence 15377 BP; 3197 A; 4630 C; 4755 G; 2774 T;

Query Match	5.58;	Score 16;	DB 1;	Length 15377;
Best Local Similarity	100.08;	Pred. No. 27;		
Matches 16;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```
QY      210 TCATTGATGCTTTGG 225
          |||||
Db      14930 TCATTGATGCTTTGG 14945
```

RESULT	6
ID	06656/c
AC	06656; standard; cDNA; 2216 BP.
DT	19-JAN-1995 (first entry)
DE	Murine delta opioïd receptor coding sequence.
KM	delta opioïd; enkephalin; receptor; mouse; murine; analgesic; pain;
KW	drug addiction; neurological disorder; psychiatric; disorder;
OS	cardiovascular disorder; ds.
MS	mus musculus.
FH	Key
FT	Location/Qualifiers
FT	cds
FT	59..1174
FT	/tag= a
FT	/product= opioïd_receptor
PN	FR697850-A.
PD	13-MAY-1984.
PF	10-NOV-1992; 013526.
PR	10-NOV-1992; FR-013526.
PA	(UYST-) UNIV PASTEUR STRASBOURG LOUIS.
PI	Kieffer B.
PI	WPI: 94-178255/22.
DR	P-PSDB: R66503.
PT	New nucleic acid encoding opioïd receptor - and related
PT	polypeptide, antisense nucleic acid, probes, recombinant cells
PT	and ligands, useful in diagnosis and treatment of e.g.
PT	neurological disorders
PS	Claim 3; Page 16-18; 29pp; French.
CC	A cDNA bank constructed from hybridoma NG108-15, was used to
CC	transfect COS-1 cells. The cells were tested for ability to bind
CC	tritium-labelled Tyr-D-Thr-Gly-Phe-Ieu-Thr. In the presence or
CC	absence of the opioïd antagonist naloxone. Clone K56 was isolated
CC	from a positive colony and found to contain a 2216bp insert. This
CC	cDNA encodes a delta opioïd (enkephalin) receptor with apparent
CC	dissociation constant 1.4nM and Bmax 3.9-6.4 pmoles/mg protein.
SC	Sequence 2216 BP; 460 A; 647 C; 649 G; 460 T;
Query Match	5.5%; Score 16; DB 1; Length 2216;
Best Local Similarity	100.0%; Pred. No. 26;
Matches 16; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	217 TGCTTTGGGCCCTT 232 
DB	2131 TGCTTTGGGCCCTT 2116
RESULT	7
ID	075927/c
AC	075927; standard; DNA; 2272 BP.
DT	17-AUG-1995 (first entry)
DE	Mouse delta opioïd receptor MOR1 cDNA.
KM	Mouse; kappa; delta; mu; opioïd receptor; brain; primer: PCR; amplify;
KW	transmembrane domain; somatostatin; Receptor; human; expression vector;
KW	truncate; Chimeric; assay; probe; ss.
OS	Mus musculus.
FH	Key
FT	Location/Qualifiers
FT	cds
FT	59..1174
FT	/tag= a
FT	/product= opioïd_receptor
PN	FR697850-A.
PD	13-MAY-1984.
PF	10-NOV-1992; 013526.
PR	10-NOV-1992; FR-013526.
PA	(UYST-) UNIV PASTEUR STRASBOURG LOUIS.
PI	Kieffer B.
PI	WPI: 94-178255/22.
DR	P-PSDB: R66503.
PT	New nucleic acid encoding opioïd receptor - and related
PT	polypeptide, antisense nucleic acid, probes, recombinant cells
PT	and ligands, useful in diagnosis and treatment of e.g.
PT	neurological disorders
PS	Claim 3; Page 16-18; 29pp; French.
CC	A cDNA bank constructed from hybridoma NG108-15, was used to
CC	transfect COS-1 cells. The cells were tested for ability to bind
CC	tritium-labelled Tyr-D-Thr-Gly-Phe-Ieu-Thr. In the presence or
CC	absence of the opioïd antagonist naloxone. Clone K56 was isolated
CC	from a positive colony and found to contain a 2216bp insert. This
CC	cDNA encodes a delta opioïd (enkephalin) receptor with apparent
CC	dissociation constant 1.4nM and Bmax 3.9-6.4 pmoles/mg protein.
SC	Sequence 2216 BP; 460 A; 647 C; 649 G; 460 T;



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:40:18 ; Search time 650.92 Seconds

(without alignments)  
111.082 Million cell updates/sec

Title: US-09-049-696-7

Perfect score: 289  
Sequence: 1 GAAATATCCACTGATGAT.....AGCTTGAGACTAAGGATTAA 289

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	6.2	1386	1 T69545	Rat pheromone rece
2	17	5.9	314	1 Q60964	Human brain Expre
3	17	5.9	1199	1 V45211	Human endosulfine
4	17	5.9	8906	1 V32979	Tn7 target plasmid
5	16	5.5	15377	1 Q25975	MH mutant porcine
6	16	5.5	2216	1 Q66566	Murine delta opiol
7	16	5.5	2272	1 Q75927	Mouse delta opiol
8	16	5.5	273	1 T21900	Human gene signatu
9	16	5.5	110000	1 T58840_0	Mycoplasma genital
10	16	5.5	2218	1 V49253	Mouse delta opiate
11	16	5.5	2981	1 V70360	Human ehb21 encodi
12	16	5.5	44453	1 X23519	Human kidney amino
13	16	5.5	1323	1 X39675	Renal cancer assoc
14	15	5.2	3905	1 N70461	Sequence encoding
15	15	5.2	3697	1 Q10324	Human Natriuretic
16	15	5.2	888	1 N60551	Mouse kidney cell
17	15	5.2	3905	1 Q25893	Human Factor XIII
18	15	5.2	107	1 Q30733	M. tuberculosis 38
19	15	5.2	106	1 Q30735	Sequence of Intron
20	15	5.2	466	1 Q43682	Sequence of the hu
21	15	5.2	5931	1 Q43701	ALL-1 gene cDNA.
22	15	5.2	14255	1 Q43526	Aspartokinase II g
23	15	5.2	2223	1 Q47809	Clostridium perfr
24	15	5.2	374	1 Q52276	11q23 chromosom
25	15	5.2	1113	1 Q62305	11q23 chromosom
26	15	5.2	1420	1 Q53475	CDNA encoding htr
27	15	5.2	4201	1 Q53477	CDNA encoding htr
28	15	5.2	11907	1 Q45339	ALL-1 (acute lymph
29	15	5.2	14255	1 Q75181	Rat hippocampal ne
30	15	5.2	1556	1 Q95035	Rat hippocampal ne
31	15	5.2	1200	1 Q95036	MIL 1.5B fragment
32	15	5.2	4201	1 T16330	MIL 1.5B fragment
33	15	5.2	4201	1 T16330	Human Serate-2 (H
34	15	5.2	4483	1 T40091	LsRNA sequence fr
35	15	5.2	669	1 T58716	LsRNA sequence fr
36	15	5.2	669	1 T58717	LsRNA sequence fr
37	15	5.2	669	1 T58718	LsRNA sequence fr
38	15	5.2	669	1 T58719	LsRNA sequence fr
39	15	5.2	669	1 T58720	LsRNA sequence fr
40	15	5.2	669	1 T58721	LsRNA sequence fr
41	15	5.2	669	1 T58722	LsRNA sequence fr
42	15	5.2	669	1 T58723	LsRNA sequence fr
43	15	5.2	669	1 T58724	LsRNA sequence fr

## ALIGNMENTS

RESULT	1	Location/Qualifiers
ID	T69545	standard; DNA; 1386 BP.
AC	T69545;	
DT	26-AUG-1997	(first entry)
DE	Rat pheromone receptor VN1 cDNA.	
KW	Pheromone receptor; vomeronasal sensory neuron; social behaviour;	
KW	maternal behaviour; reproductive behaviour; fertility;	
KW	hormone secretion; ss.	
OS	Rattus sp.	
FH	Key	
FT	cds	270..1217
FT		/*tag- a
PN	W09714790-A1.	
PD	24-APR-1997.	
PF	18-OCT-1996; U16637.	
PR	19-OCT-1995; US-005698.	
PA	(YXCO ) UNIV COLUMBIA NEW YORK.	
PI	Kael R, Dulac G;	
DR	WPI; 97-245107/22.	
DR	P-PSDB; W19103.	
PT	Nucleic acid molecule encoding vertebrate pheromone receptor -	
PT	useful to identify modulators for control of reproductive and social	
PT	behaviour, fertility and hormone secretion	
PS	Claim 1; Fig 8; 123pp; English.	
CC	A cDNA clone (T69545) codes for rat pheromone receptor VN1	
CC	(W19103), a member of a novel family of presumed 7-transmembrane	
CC	domain receptors that are evolutionarily independent of the odorant	
CC	receptors of the main olfactory epithelium (MOE). It was isolated	
CC	by differential screening of cDNA libraries constructed from rat	
CC	single vomeronasal sensory neurons. A VN1 insert in pBluescript	
CC	has been deposited as ATCC 97294. VN1 is expressed only in about	
CC	4% of VSN and never in the MOE. cDNA clones for VN2-7 were obt.	
CC	by PCR and hybridisation (see also T69546-50, VN2 sequence not	
CC	given). VN polypeptides (W19103-09) can be expressed in host-	
CC	vector systems for use in identifying modulators for control of	
CC	maternal, reproductive and social behavior, to increase fertility,	
CC	control hormone secretion and to regulate food uptake in humans and	
CC	animals. A human homologue, HG25 (T69551), has also been obt.	
SQ	Sequence 1386 BP; 357 A; 322 C; 271 G; 435 T;	

Query Match 6.2%; Score 18; DB 1; Length 1386;

Best Local Similarity 100.0%; Pred. No. 2.3;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	125 TCTGCAGCTCAAGACTA 142
DB	1042 TCTGCAGCTCAAGACTA 1059
RESULT	2
ID	Q60964
AC	Q60964;
DT	16-MAR-1994 (first entry)
DE	Human brain Expressed Sequence Tag EST01042.
KW	Gene transcription product; genetic markers; tagging; in vivo;
KW	transcription; mapping; locations; chromosomes; chromosomal; ss.
OS	Homo sapiens.
PN	W09316178-A.
PD	19-AUG-1993.
PF	12-FEB-1993; U01294.
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI	Adams MD, Moreno RF, Venter CJ;

LsRNA sequence fr  
Enterococcus faeca



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:32:50 ; Search time 3075.15 Seconds  
(without alignments)  
141.117 Million cell updates/sec

Title: US-09-049-696-5

Perfect score: 220  
Sequence: 1 CTAAAGTTGAATTCTGTACA.....GACAAATCTGAGAGCATGCGC 220

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est23:\*  
55: em\_est24:\*  
56: em\_est25:\*  
57: em\_est26:\*  
58: em\_est27:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	220	100.0	480	35	AA581198	AA581198 nd38c07.r
2	125.4	57.0	427	34	AA498535	AA498535 vH40f12.r
3	58.2	26.5	372	38	AA755896	AA755896 vU04f06.r
4	36	15.4	435	35	AA584450	AA584450 n005g01.s
5	34.8	15.8	285	27	AA032127	AA032127 RRAMCA211
6	34.4	15.6	453	20	D27576	D27576 CELK014EXR
7	33	15.0	327	23	F15082	F15082 SSC6A10 Por
8	33	15.0	288	33	AA450336	AA450336 zX04q12.s
9	33	15.0	433	37	AA676984	AA676984 z154b01.s
10	32.2	14.6	443	36	AA616881	AA616881 SMOVL3CA3
11	32.2	14.6	436	45	A1362967	A1362967 qy87g12.x
12	31.8	14.5	404	24	N23699	N23699 yw46c03..s1
13	31.8	14.5	263	27	C06194	C06194 C06194 Huma
14	31.8	14.5	590	36	C79364	C79364 C79364 Mous
15	31.8	14.5	668	43	A1162429	A1162429 A017P40u
16	31.8	14.5	317	49	AV008533	AV008533 AV008533
17	31.8	14.5	311	49	AV018065	AV018065 AV018065
18	31.6	14.4	281	21	R07421	R07421 y96h08..r1
19	31.6	14.4	414	39	AA828323	AA828323 ob55f09.s
20	31.6	14.4	149	43	A1179739	A1179739 EST223467
21	31.4	14.3	361	39	AA873934	AA873934 vW85f03.r
22	31.4	14.3	426	44	A1248483	A1248483 qh71d11.x
23	31.2	14.2	402	41	A1071450	A1071450 UT-R-Cl-K
24	31.2	14.2	484	46	A1422242	A1422242 Lf58h08.x
25	31	14.1	417	27	AA008223	AA008223 mg70c12.r
26	31	14.1	669	27	AA013624	AA013624 mh10f06.r
27	31	14.1	501	27	AA050085	AA050085 m108d07.r
28	31	14.1	180	28	AA114762	AA114762 m14b09.r
29	31	14.1	455	29	AA135889	AA135889 z122c12.r
30	31	14.1	440	44	AV037076	AV037076 AV037076
31	31	14.1	276	48	A1550139	A1550139 m14b09.y
32	31	14.1	627	48	A1553369	A1553369 v140c11.y
33	31	14.1	466	48	A1601422	A1601422 fc11a09.x
34	30.8	14.0	415	38	AA737000	AA737000 nx89c05.s
35	30.8	14.0	383	39	AA872500	AA872500 O110e01.s
36	30.8	14.0	429	49	A1643115	A1643115 mm54c06.y
37	30.8	14.0	429	50	A1638404	A1638404 fc38d04.x
38	30.6	13.9	268	21	R09997	R09997 yf34e04..r1
39	30.6	13.9	385	23	H18883	H18883 yN52d09..r1
40	30.6	13.9	378	23	H29932	H29932 yN81f03..r1
41	30.6	13.9	395	30	AA241400	AA241400 mw17f09.r
42	30.6	13.9	366	30	AA241407	AA241407 mw17h09.r
43	30.6	13.9	348	42	A1105738	A1105738 SWAMCA25
44	30.6	13.9	441	45	A1359809	A1359809 qy46d05.x
45	30.6	13.9	536	46	A1461240	A1461240 sa60e07.y

#### ALIGNMENTS

RESULT 1  
AA581198  
LOCUS AA581198 480 bp mRNA  
DEFINITION nd38c07.r1 NCI-CCAP\_Col1 Homo sapiens cDNA clone IMAGE:802572 5',  
similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL.  
ACCESSION AA581198  
NTD 92358970

```

VERSION      AA581198.1  GI:2358970
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 480)
AUTHORS      NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
             Unpublished (1997)
JOURNAL      On May 9, 1995 this sequence version replaced gi:803001.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.L.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michel
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies Inc., David Krizman,
Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrrp/image/image.html

Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 405.
Location/Qualifiers
1..480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:802572"
/clone_lib="NCI-CCAP_Col1"
/tissue_type="Pulk tumor"
/lab_host="IDH108"
/notes="Organ: colon; Vector: pCMV-Sport2; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT."
BASE COUNT  168 a 101 c 97 g 114 t
ORIGIN

Query Match          100.0%; Score 220; DB 35; Length 480:
Best Local Similarity 100.0%; Pred. No. 1.7e-50;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTATAGTTGAATTCCTGTCACAGAACAAACCAACAAAGAGAGCTCCAAACAGCAAAATC 60
DB 215 CTATAGTTGAATTCCTGTCACAGAACAAACCAACAAAGAGAGCTCCAAACAGCAAAATC 274

OY 61 AAAAAATGCAATCTCCGAGACACATGGGAAGTATCCGTGATCTCTAGAGCACTTAAGAAA 120
DB 275 AAAAAATGCAATCTCCGAGACACATGGGAAGTATCCGTGATCTCTAGAGCACTTAAGAAA 334

OY 121 CCACCTCATATGACACACACAGCCACCAAAATCCACCTTCATTTGCTGGAGATTGGACAAA 180
DB 335 CCACCTCATATGACACACACAGCCACCAAAATCCACCTTCATTTGCTGGAGATTGGACAAA 394

OY 181 GAATTGTGTGTTAGTCTTGACAATATCTGGAAGCATGGC 220
DB 395 GAATTGTGTGTTAGTCTTGACAATATCTGGAAGCATGGC 434

RESULT 2
LOCUS       AA498535
DEFINITION  vha0f12.i1 Barstead mouse pooled organs MPRB4 Mus musculus cDNA
            clone IMAGE:889487 5' similar to TR:G1184066 G1184066
            CALCIUM-ACTIVATED CHLORIDE CHANNEL. ;, mRNA sequence.
ACCESSION   AA498535
VERSION     92233558
KEYWORDS    EST.
            GI:2233558

```

SOURCE	house mouse.
ORGANISM	Mus musculus
TITLE	Eutheria; Rodentia; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE	1 (bases 1 to 427)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kibcad,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
JOURNAL	The WashU-HM Mouse EST Project
COMMENT	Unpublished (1996) On Sep 12, 1996 this sequence version replaced gl.1294625.
FEATURES	Contact: Marra M/Mouse EST Project WashU-HM Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:517447 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 377. Location/Qualifiers . . 427 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /map="13" /clone="IMAGE:889487" /clone_id="Barstead mouse pooled organs MPLRB4" /sex="mixed" /issue_type="pooled organs" /dev_stage="7 day" /lab_host="DH10B" /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGATCATCGAAGTCGGAGCGGCCGCTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTGTAATGGTGAC] , digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bob Barstead."
BASE COUNT	143 a 108 c 83 g 93 t
ORIGIN	
Query Match	57.0%; Score 125.4; DB 34; Length 427;
Best Local Similarity	82.3%; Pred. No. 9.7e25;
Matches 144; Conservative	0; Mismatches 31; Indels 0; Gaps 0;
Dy	1 CTATAGTGAATTCTGTGACAGAACAACCAACAAAGAGCTCCAAACAGCAAAATC 60
Dy	68 CTGTGTTGATTTCTGTACAGAAAAAATTCACAAACAGAACCCCAATATGACAAAAC 127
Dy	61 AAAAAATGCATCTCCGAAGACACATGGGAGTAGTCCGTGATTTCTGAGGACTTTAAGAAA 120
Dy	128 AACGATGCAATCTCCGAGACACGTGGGTAAGTATCCAGGAATCTGAGGACTTCAAGCAA 187
Dy	121 CCACGCCATATACCAACAGCACCAACAAATCCCACCTTCTCAATTTGTCGAGATGG 175
Dy	188 CCACTCCATACAGACCACCACTGCACCACTTCTCACTGTTGCCAAAATGG 242
RESULT 3	
LOCUS	AA755896 372 bp mRNA EST 21-JAN-1998
DEFINITION	v004f06.f1 Soares mouse mammary gland Nbhmg Mus musculus cDNA clone
IMAGE:1179683 5	similar to SW:EBLC_BOVIN P54281 EPITHELIAL CHLORIDE CHANNEL PROTEIN ; , mRNA sequence.

ACCESSION AA755896  
NTD 92803094  
VERSION AA755896.1 GI:2803094  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 372)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
TITLE The WashU-HMT Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:2044941.  
Contact: Marra M/Mouse EST Project  
WashU-HMT Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:637531  
Seq primer: -28m13 rev2 ET from Amer sham  
High quality sequence stop: 365.  
Location/Qualifiers  
1..372  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/map="X: 2: 726H05: 12"  
/clone="IMAGE:1179683"  
/clone\_lib="Soares mouse mammary gland NBMWG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="organ: mammary gland; Vector: pT773D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15' TGTTACCAATCGAAGTGGAGCGCGCGCGGAGTGTGTTTTTTTTTTTTTTT T 3'; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonafido."  
BASE COUNT 121 a 86 c 81 g 84 t  
ORIGIN  
Query Match 26.5%; Score 58.2; DB 38; Length 372;  
Best Local Similarity 64.2%; Pred. No. 1.9e-06;  
Matches 106; Conservative 0; Mismatches 53; Indels 6; Gaps 1;  
QY 61 AAAATGCAATTCGAGACATGGAATGATCCGTGATTTGAGAGACTTAAAGAAA 120  
DB 2 AATGTGCAATCGACAGACAGCCTGGATGTAATCAAGAGCTGCTGATTTTCAAGATG 61  
QY 121 CCACCTCTATG-----ACACAGACGACCAATCCACCTCTATGCTGACAGTTTG 174  
DB 62 CCCCTCCATGAGAGAGAGAGCCCTCTCCACCTAATTTTCAAGTCCA 121  
QY 175 GACAAAGAATTTGTGTGTTAGTCTTACAAATCTGGAAGCATGG 219  
DB 122 GAAGGGAGTGTGTGTTGTCAGATTAATCTGGAAGCAAG 166

RESULT 4  
AA584450/c 435 bp mRNA EST 26-SEP-1997  
LOCUS no05g01.s1 NCI-CGAP\_Phel Homo sapiens cDNA clone IMAGE:1099824 3'  
DEFINITION similar to contains element L1 repetitive element ;, mRNA sequence.  
ACCESSION AA584450  
NTD 92369059  
VERSION AA584450.1 GI:2369059  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 435)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1290624.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Stratagene, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.Dio.llnl.gov/dbp/image/image.html  
Insert Length: 1532 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amer sham  
High quality sequence stop: 306.  
Location/Qualifiers  
1..435  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1099824"  
/clone\_lib="NCI-CGAP\_Phel1"  
/tissue\_type="pheochromocytoma"  
/lab\_host="SOLR (Ranaycin resistant)"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT.  
Pheochromocytoma. 5' adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'  
Average insert size: 1.3 kb."  
BASE COUNT 72 a 110 c 109 g 144 t  
ORIGIN  
Query Match 16.4%; Score 36; DB 35; Length 435;  
Best Local Similarity 58.3%; Pred. No. 2.2;  
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
QY 18 ACAGACAAACACACACAGAGAGCTCCAAACCAATCAATATGCAATCTCGCA 77  
DB 276 AAGAGAAAGAAAAAGAAAAAGAACACAAAAAACAATTAATCTTTAGCCTGA 217  
QY 78 AGCAGTGGAGAGATCCGTGATTTGAGAGACTTTAAGAAAAACACT 125  
DB 216 TCGTCAAGGGCAGAGAGATTTAGAGCTGGGCTTGTGTGGCCCT 169  
RESULT 5  
AA032127 285 bp mRNA EST 12-DEC-1996  
LOCUS RRACCA211SR Brugia malayi adult male cDNA (SAM94NL-BRM) Brugia  
DEFINITION malayi cDNA clone RRACCA211 5', mRNA sequence.  
ACCESSION AA032127  
NTD 91502099



BASE COUNT	78 a	82 c	91 g	72 t	4 others
ORIGIN	/clone.lib="Porcine small intestine cDNA library" /note="directionally cloned cDNA in XL1-blue MRF" "				

Query Match	15.0%;	Score 33;	DB 23;	Length 327;
Best Local Similarity	87.8%;	Pred. No. 14;		
Matches 36;	Conservative	0;	Mismatches 5;	Indels 0;
			Gaps	0

Dy 180 AGAATTGTGTTTACCTCCTTGACAATCTGGAGCATGCC 220  
||||| ||||| ||||| ||||| ||||| |||||  
Db 1 AGAATCGTGTCTCACTCCCTGCACAAGTCGGAGCATGAC 41

RESULT	8
AA450336/c	
LOCUS	
DEFINITION	AA450336 288 bp mRNA EST 04-JUN-1997
IMAGE:785542_3'	zx04q12.s1 Soares.totai.fetus_Nb2HF8_9w Homo sapiens cDNA clone
IMAGE:785542_3'	IMAGE:785542_3' mRNA sequence.

ACCESSION	AA450336
NID	g2162881
VERSION	AA450336.1
KEYWORDS	GI:2162881
SOURCE	EST.
	human.

ORGANISM	REFERENCE	AUTHORS
<i>Homo sapiens</i>	1 (bases 1 to 288)	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
<i>Eukaryota</i> ; <i>Metazoa</i> ; <i>Chordata</i> ; <i>Cranialia</i> ; <i>Vertebrata</i> ; <i>Mammalia</i> ;		
<i>Eutheria</i> ; <i>Primates</i> ; <i>Catarrhini</i> ; <i>Hominae</i> ; <i>Homo</i> .		

TITLE  
 JOURNAL  
 COMMENT  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced g1:1407011

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: -41ml3 fwd. Et from Amersham  
High quality sequence stop: 258.

FEATURES	Location/Qualifiers
source	1. . 288

```

/organism="Homo sapiens"
/db_xref="GDB:5983160"
/db_xref="taxon:9606"
/map="952605; 1; 1q32.1-1q32.3"
/clone="IMAGE:785542"
/clone.lib=" Soares.total_fetus_ND2HF8_9w"
/dev_stage="8-9 weeks"
/ld_host="DH10B"
/node_vector: pf773d-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGGAGGAGCGCCGCTAAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pf773 vector. Library
was through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

```

BASE COUNT	72 a	56 c	59 g	101 t
ORIGIN				

Query Match	15.0%;	Score 33;	DB 33;	Length 288
Best Local Similarity	69.2%;	Pred. No. 14;		

Matches	45	Conservative	0	Mismatches	20	Indels	0	Gaps	0
QY	6	GTTGATTCGTGACAGACAAACCACAAGAGCTCAACACAGCAAAATCAAAA	65						
Db	66	GTTTAATATGTAAGAAAAACAAACACACAAACAAAAA	7						

QY	66	TGCA	7
Db	6	AAAA	2

RESULT	9				
AA676984/c					
LOCUS	AA676984	433 bp	mRNA		
DEFINITION	zj54b01.s1 Soares.fetal.liver.spleen.INTLS.S1 Homo sapiens cDNA				
ACCESSION	AA676984				
	clone IMAGE:454057 3', mRNA sequence.				
	AA676984				
				EST	19-DEC-1997

ACCESSION	AA070504	
NID	g2657506	
VERSION	AA676984.1	GI:2657506
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM	
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia	
Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	
1 (bases 1 to 433)	
AUTHORS	
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Getse,J.G., Jost	

TITLE  
JOURNAL  
COMMENT

Washin-NCI human EST Project  
Unpublished (1997)  
On May 5, 1995 this sequence version replaced gi:798292.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estewatson.wustl.edu](mailto:estewatson.wustl.edu)  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 424.

## FEATURES

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19"
/clone="IMAGE:454057"
/clone_1fb="Soares_fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen Vector: p773D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI
This is a subcloned version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'.
AAGCGAGAGATTTATTTATTAAGATCTTTTTTTTTTTTTTTTTTT 3'] ,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia) , digested with Pac I and cloned into the Pac
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```

	a	c	g	t	others
BASE COUNT	136	54	52	186	5
ORIGIN					

Query Match	15.0%;	Score 33;	DB 37;	Length 433;
Best Local Similarity	54.5%;	Pred. NO. 15;		
Matches 60;	Conservative 0;	Mismatches 50;	Indels 0;	Gaps 0;

9 GAATTCGTACAGAACAAACCACACAAAGAAGCTCCAAACAAGCAAATCAAAAATGC 68

Db 339 GAACATGCAAAAAAATAAANNCCAAANNAAAAA 280

Qy 69 AATCTCCGAGCAGTGGAGTGCCTGATTCGAGACTTTAGAA 118  
LOCUS 118  
DEFINITION Onchocerca volvulus infective larva CDNA  
(SAM94WL-OVL3) Onchocerca volvulus CDNA clone SWJIC039 5', mRNA  
sequence.

RESULT 10  
AA618681 443 bp mRNA EST 12-NOV-1997  
LOCUS AA618681  
DEFINITION Onchocerca volvulus infective larva CDNA  
(SAM94WL-OVL3) Onchocerca volvulus CDNA clone SWJIC039 5', mRNA  
sequence.

ACCESSION  
NID 92522565  
VERSION AA618681.1 GI:2522565  
KEYWORDS  
SOURCE EST.  
ORGANISM Onchocerca volvulus.  
Onchocerca volvulus.  
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spiruridae;  
Filarioidea; Onchocercidae; Onchocerca.  
REFERENCE  
AUTHORS Williams, S.A., Lu, W., Lizotte-Manlewski, M. and Laney, S.J.  
TITLE Genes expressed in infective third stage larvae of Onchocerca  
volvulus  
JOURNAL Unpublished (1995)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404802.

FEATURES  
SOURCE  
Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genomesmith.edu  
Seq primer: pbluescript SK.  
Location/Qualifiers  
1. 443  
/organism="Onchocerca volvulus"  
/strain="Sierra Leone"  
/db\_xref="taxon:6282"  
/clone="SWJIC0399"  
/clone\_lib="Onchocerca volvulus infective larva CDNA  
(SAM94WL-OVL3)"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: lambda Unizap XR; Site 1: EcoR I; Site 2:  
Xho I; Cutaneous filarial nematode parasite of humans  
RNA was prepared from third stage infective larvae of  
Onchocerca volvulus isolated from mosquitoes 10 days after  
infection and converted to double stranded CDNA using  
reverse transcriptase and oligo(dt) followed by Rnase H  
and DNase I. The library had 1.8 x 10E5 independent  
recombinants and average insert size was 900 base pairs.  
The library was constructed by Weihong Lu. The library is  
available from Dr. S.A. Williams, email genomesmith.edu."

BASE COUNT 74 a 45 c 90 g 206 t 28 others

ORIGIN

Query Match 14.6%; Score 32.2; DB 36; Length 443;  
Best Local Similarity 61.2%; Pred. No. 24;  
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 4 TAGTGAATTCGTACACAAACCAACCAAGAGCTCCAAAGCAATCAAA 63  
Db 224 TAGTGAATTCGTACACAAACCAACCAAGAGCTCCAAAGCAATCAAA 175

Qy 64 AATGCAATCTCCGAGCACA 83  
Db 174 AAAGTAAACCAAGAGCAAAA 155

RESULT 11  
A1362967 436 bp mRNA EST 16-FEB-1999  
LOCUS A1362967  
DEFINITION qv87g12.x1 NCI-CGAP\_Brn25 Homo sapiens CDNA clone IMAGE:2019046 3',  
mRNA sequence.

ACCESSION  
NID 94114588  
VERSION A1362967.1 GI:4114588  
KEYWORDS  
SOURCE EST.  
ORGANISM human.  
human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTCAP), Tumor Gene Index  
(CGAP/BTCAP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT On May 8, 1995 this sequence version replaced gi:801270.

FEATURES  
SOURCE  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.llnl.gov/bdrrp/image/image.html  
Insert Length: 1669 Std Error: 0.00  
Seq primer: -400P from GIBCO  
High quality sequence stop: 378.  
Location/Qualifiers  
1. 436  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2019046"  
/clone\_lib="NCI-CGAP\_Brn25"  
/issue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pTZ19-Pac (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand CDNA was primed with a Not I - oligo(dt) primer [5'  
TGTTACCAATCTGAGAGGAGCGCGCATAGCTTTTCTTTTCTTTTCTTTT  
T 3']; double-stranded CDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pTZ19 vector.  
Library is normalized, and was constructed by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 179 a 76 c 98 g 83 t

ORIGIN

Query Match 14.6%; Score 32.2; DB 45; Length 436;  
Best Local Similarity 57.4%; Pred. No. 24;  
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 20 AGAACAACCAACCAACCAAGAGCTCCAAACCAAGCAATCAAAATGCAATCCGAG 79  
Db 117 AGAACAACCAACCAACCAAGAGCTCCAAACCAAGCAATCAAAATGCAATCCGAG 176

Qy 80 CACATGGGAAGTGTCCGTGATTCGTGAGCACTTTAGAAA 120  
Db 177 TACTTTGGAAGGAGAGACTTGTGATGATTCACGAAA 217

RESULT 12

N23699 404 bp mRNA EST 28-DEC-1995  
LOCUS yw46c03.s1 weizmann Olfactory Epithelium Homo sapiens cDNA clone  
DEFINITION IMAGE:255268 3' similar to gb:X81002 RETROVIRUS-RELATED POLYPROTEIN (HUMAN); contains Alu repetitive element; contains element L1 repetitive element ;, mRNA sequence.  
N23699  
ACCESSION g1137849  
NID N23699.1 GI:1137849  
VERSION EST.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 404)  
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapell,B., Chisoso,S., Dietrich,N., Dubuque,T., Faveillo,A., Gish,M., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasz,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.  
TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT On Apr 14, 1993 this sequence version replaced gi:693148.  
TITLE Wilson RK  
JOURNAL Washington University School of Medicine  
MEDLINE 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
COMMENT Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
High quality sequence stops: 363  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: Polyt not found  
Insert Length: 1895 Std Error: 0.00  
Seq primer: m13 -40 forward  
High quality sequence stop: 363.  
Location/Qualifiers  
1..404  
/organism="Homo sapiens"  
/db\_xref="GDB:3865677"  
/db\_xref="taxon:9606"  
/map="12: 982C08: 17: 957F08: 930D05: 17q11.2-q23: 17q21.1-17q22: 17q21.1-17q22"  
/clone="IMAGE:255268"  
/clone\_lib="Weizmann Olfactory Epithelium"  
/sex="Female"  
/tissue\_type="olfactory epithelium"  
/dev\_stage="35 year old"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: nose. Vector: pBluescript SK-. Site1: EcoRI; Site2: XhoI; Cloned unidirectionally. Primer: 0.1kb 5' Olfactory epithelium, normal. Average insert size: 0.8 kb; Uni-ZAP XR Vector. Library constructed by N. Walker, D. Lancelot, Weizmann Institute of Science. -5' adaptor sequence: 5' GAATTCGGCGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

BASE COUNT 157 a 75 c 97 g 69 t 6 others  
ORIGIN

Query Match 14.5%: Score 31.8; DB 24; Length 404;  
Best Local Similarity 67.2%: Pred. No. 31;  
Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Oy 26 AACCAACAAGAGCTCCAAACAGCAATCAAAATCAATCTCCGAGCAGCATG 85  
Db 158 AACCAAGAAGAGAGATCCAAATAGCTCAATTAAGAAATGAATGGAGATGGCCAG 217

Oy 86 GGAAGC 92  
Db 218 CCGAGTG 224

RESULT 13  
LOCUS C06194 263 bp mRNA EST 16-OCT-1996  
DEFINITION C06194 Human pancreatic islet Homo sapiens cDNA clone hbc5652  
ACCESSION C06194  
NID g1502970  
VERSION C06194.1 GI:1502970  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 263)  
AUTHORS Takeda,J.  
TITLE Human pancreatic islet ESTs  
JOURNAL Unpublished (1995)  
COMMENT On May 9, 1995 this sequence version replaced gi:802384.  
TITLE Jun Takeda  
JOURNAL Institute for Molecular and Cellular Regulation, Gunma University  
MEDLINE 3-39-15 Showa-machi, Maebashi Gunma 371, Japan  
COMMENT Tel: 272-20-8856  
Fax: 272-20-8896  
Email: jtakeda@sh.gunma-u.ac.jp.  
Location/Qualifiers  
1..263  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="2"  
/clone="hbc5652"  
/clone\_lib="Human pancreatic islet"  
/note="Vector: Lambda ZapII; Site1: Eco RI; Site2: Xho I; mRNA was prepared from normal adult human islets. cDNA was directionally synthesized from the Xho I in the vector to the EcoRI site. cDNA was size fractionated to remove sequences <1000 bp in size."

FEATURES  
SOURCE  
1..263  
Location/Qualifiers

BASE COUNT 92 a 38 c 45 g 88 t  
ORIGIN

Query Match 14.5%: Score 31.8; DB 27; Length 263;  
Best Local Similarity 48.1%: Pred. No. 29;  
Matches 90; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Oy 10 AATCTGTACAGACAAACCAACCAAGAGCTCAACACAGCAAAATCAAAATGCA 69  
Db 51 AATTTGTGGCATTAATAATTAGAGCATTAAGCTGATGATATATACCAAAACGAG 110  
Oy 70 ATCTCCGAAGCAGATGGAGATGATCCGTATCTGAGAGCTTTAAGAAACCATCTCCTA 129  
Db 111 AATTAGATTACCAAAATATGGGAATAGAGTGTTCATTAATTCAGATTCAATCAAT 170  
Oy 130 TCACAACAGCAGCAACCAATCCACCTCTCATCTGACATTTGACAAAGAAATTTGT 189  
Db 171 TTAGGAACAAACCAACCCATGCTTACATTGATTTGTCCACAGATTAACTTCAGTAGTGC 230  
Oy 190 GTTAGT 196  
Db 231 ATTAGT 237

RESULT 14  
LOCUS C79364 590 bp mRNA EST 26-JUN-1998  
DEFINITION C79364 Mouse 3.5-dpc blastocyst cDNA Mus musculus cDNA clone  
ACCESSION J006507 3', mRNA sequence.  
C79364





OM nucleic - nucleic search, using sw model

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

Run on: August 6, 1999, 04:02:32 ; Search time 921.96 Seconds

(without alignments)  
68.385 Million cell updates/sec

Title: us-09-049-696-6

Perfect score: 252

Sequence: 1 CAAGCAATGTGTGTTAGT.....GAGGACGTCATCTGCAGC 252

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N.Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30.8	12.2	1455	T96843	Intron 13 of human
2	28.4	11.3	4478	Q13363	HER-3 epithelial g
3	28.2	11.2	3581	Q44391	Sequence of murine
4	28.2	11.2	40352	V02032	MAGE-B cluster DNA
5	27.8	11.0	5211	Q77789	Pro-pro-cobra C3 c
6	27.8	11.0	5924	Q77790	Cobra CVPI coding
7	27.8	11.0	4138	Q77791	Cobra partial CYP2
8	27.6	11.0	2890	V66832	Zea mays soluble s
9	27.6	11.0	8655	V69750	Mus dunni endogene
10	27.6	11.0	343	X51918	Human secreted pro
11	27.4	10.9	1130	Q50187	Porcine Factor VII
12	27.4	10.9	1623	Q66516	Sequence of porcine
13	27.4	10.9	1130	V12114	Sus scrofa factor
14	27.4	10.9	6402	V12113	Sus scrofa factor
15	27.4	10.9	4334	V12142	Homo sapiens facto
16	27.4	10.9	1130	V25811	Porcine factor VII
17	27.4	10.9	4700	X26908	4.7 kb transcript
18	27.2	10.8	999	T90124	Human CRF1 (TRAF-
19	27.2	10.8	1445	V44294	Human secreted pro
20	27	10.7	995	Q10491	Cationic granule p
21	27	10.7	907	T05128	Human heparin-bind
22	27	10.7	772	T05129	Human heparin-bind
23	27	10.7	262	T65697	Polymorphic repeat
24	26.8	10.6	3500	N90229	Mouse tumour necro
25	26.8	10.6	492	O51039	Human glucokinase
26	26.8	10.6	492	O51040	Human glucokinase
27	26.8	10.6	4138	O55179	Human orphan recep
28	26.8	10.6	4905	T10059	Human orpnan recep
29	26.8	10.6	479	V52734	Human hepatocyte n
30	26.8	10.6	6260	V52729	Human hepatocyte n
31	26.8	10.6	4905	V58734	ErbB-3 glycoprotei
32	26.6	10.6	3254	Q30659	Eph-related PTK Ce
33	26.6	10.6	4451	V23288	Synthetic human Fa
34	26.4	10.5	5406	N93197	Fragment of clone
35	26.4	10.5	2640	N80220	Sequence of Mycoba
36	26.4	10.5	7076	O20186	Hiya gene with ctx
37	26.4	10.5	5438	Q29263	Human calcitonin chan
38	26.4	10.5	7076	O86906	Vibrio cholerae di
39	26.4	10.5	278	T65703	Polymorphic repeat
40	26.4	10.5	1752	T92237	Mercuric ion reduc
41	26.4	10.5	1752	T92238	Mercuric ion reduc
42	26.4	10.5	1746	T92239	Mercuric ion reduc
43	26.4	10.5	1752	T92240	Mercuric ion reduc

## ALIGNMENTS

44 26.4 10.5 1752 1 T92241  
45 26.4 10.5 1232 1 V32416

Mercuric ion reduc  
Homo sapiens clone

RESULT 1

ID T96843 standard; DNA; 1455 BP.

AC T96843:

DE 27-APR-1998 (first entry)

DE Intron 13 of human pRb2/p130 tumour suppressor gene.

KW Retinoblastoma susceptibility gene; pRb2; p130; pRb2/p130 gene;

KW cell cycle; tumour suppressor gene; cancer; molecular marker;

KW diagnosis; prognosis; predisposition; endometrial carcinoma;

KW ovary cancer; lung squamous cell carcinoma; lung adenocarcinoma;

KW human; intron 13; ds.

OS Homo sapiens.

PN W09738125-A1.

PD 16-OCT-1997.

PF 03-APR-1997; U05598.

PR 03-MAR-1997; US-039532.

PR 05-APR-1996; US-014943.

PR 05-JUN-1996; US-019372.

PR 21-JUN-1996; US-020196.

PA (UVE-) UNIV JEFFERSON THOMAS.

PI Baldi A, Giordano A;

PI WPI: 97-512731/47.

PT Tumour suppressor pRb2/p130 gene intron and promoter sequences -

PT pre-disposition to cancer and prognosis of cancer and predicting

PS Clam 63; Page 113; 169pp; English.

CC This genomic DNA sequence comprises intron 13 of the human

CC tumour suppressor pRb2/p130 gene that plays a key role in cell

CC cycle progression. The gene was isolated from a human pl genomic

CC library using primers (see T96897-98) based on the published cDNA

CC sequence, with the 5' flanking region obtained from a placenta

CC genomic library. The expression level of the pRb2/p130 gene in a

CC sample from a patient can be determined and used to: (1) determine

CC a prognosis in a patient afflicted with cancer; (ii) detect a

CC cancerous disease state in a tissue; (iii) identify individuals at

CC risk for cancer, or individuals at risk for the recurrence of

CC cancer after treatment; or (iv) grade a cancer: where the level of

```

ID 013363 standard; cDNA; 4978 BP.
AC 013363;
DE 13-NOV-1991 (first entry)
DE HER-3 epithelial growth factor homologue gene.
KW Receptor tyrosine kinase; EGFR; HER2; neoplasia; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 199..4227
FT signal_peptide 199..255
FT mat_peptide 256..4227
FT misc_rna 2095..2190
FT /*tag= d
FT /*note= "transmembrane region"
PD EP-444961-A.
PD 04-SEP-1991.
PF 01-MAR-1991; 301737.
PF 02-MAR-1990; US-488347.
PA (BRIM) BRISTOL-MYERS SQUIB.
PI Plozman GD, Shoyab M, Whitney GS;
DR WPI: 91-261699/36.
DR P-PSDB: R13833.
PT New epithelial growth factor receptor homologue HER3 - used to
PT diagnose and treat conditions involving HER3, EGFR and HER2-neu
PT e.g. neoplastic diseases.
PS Claim 1: Fig 1: 20pp; English.
CC The sequence encodes an HER3 precursor and can be obt'd. from cell
CC sources which produce HER3-like activities and/or express HER3
CC encoding mRNA, e.g. epidermoid and breast carcinoma cell lines, or
CC normal colon, kidney and brain. The HER3 DNA can be used to detect
CC and quantify mRNAs encoding HER3. HER3 mRNA levels may indicate
CC emerging and/or existing neoplasias as well as the onset and/or
CC progression of other human diseases. Anti-sense HER3 RNA mols. may
CC be useful therapeutically to inhibit the translation of HER3
CC encoding mRNAs to eliminate the presence of HER3. They may also be
CC useful in elucidating HER3 functional mechanisms.
CC Sequence 4978 BP; 1194 A; 1315 C; 1321 G; 1148 T;
SQ

```

Query Match 11.3%; Score 28.4; DB 1; Length 4978;  
 Best Local Similarity 54.9%; Pred. No. 7.4;  
 Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```

QY 130 GTGACATTGACAGTCTGCCATGTACAAAGTGAAGTACATGATTAAGAGTGGCGCT 189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2558 GTGACAAAGCTGACAGAGTACACCTGGCATAGTCCACAGACCTTACAAATGGGCGATGG 2499
QY 190 GACAGGACACACTGCGCCAAAGATTACCTGCAGCAGCTTCA 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2498 TCCAGGGTGCCCAATGGCCGATATGATCTGTACACAGCTTGA 2457

```

RESULT 3  
 044391  
 ID 044391 standard; cDNA to mRNA; 3581 BP.  
 AC 044391;  
 DT 14-SEP-1994 (first entry)  
 DE Sequence of murine OSF-4 cDNA.  
 DE OSF-4; cadherin; growth factor; osteogenesis; osteoblast; therapy;  
 KW diagnosis; ss.  
 OS Mus musculus.  
 FH Key Location/Qualifiers  
 FT cds 284..2671  
 FT /\*tag= a  
 PD EP-585801-A.  
 PD 09-MAR-1994.  
 PF 25-AUG-1993; 113602.  
 PF 28-AUG-1992; JP-230028.  
 PA (FARH) HOECHST JAPAN LTD.  
 PI Amano E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;  
 WPI: 94-076152/10.

```

DR P-PSDB: R49730.
PT New bone related, cadherin-like OSF-4 proteins - for treatment
PT and diagnosis of bone metabolic disease, and nucleic acid
PT encoding them
PS Claim 3: Page 13-17, 34pp; English.
CC cDNA libraries were constructed from the mouse osteoblastic cell
CC line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
CC then as much common DNA as possible removed by hybridisation between
CC the 2 libraries. Residual E1-specific DNA was amplified, inserted
CC into lambda gt10 and screened by plaque hybridisation. A minibank of
CC 273 E-specific clones was recovered, their inserts amplified and
CC used to screen total RNA from both cell types. One clone specific
CC for E1 was identified and sequenced. The insert from this clone was
CC cloned to screen cDNA prepd. from E1 RNA and the longest posn. insert
CC used in pGEM 112f (+) to give pK0164. This insert was sequenced;
CC it encoded the 795 AA mouse precursor protein (044391/R49730). The
CC insert was also used to screen a cDNA bank prepd. from human
CC osteosarcoma to identify 2 clones encoding the 2 human precursor
CC proteins - OSF-4-1 and OSF-4-2 (044392/R49731 and 044393/R49732
CC respectively).
SQ Sequence 3581 BP; 1121 A; 774 C; 800 G; 886 T;

```

Query Match 11.2%; Score 28.2; DB 1; Length 3581;  
 Best Local Similarity 54.3%; Pred. No. 7.5;  
 Matches 57; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```

QY 145 GCTGCCATGTACAAAGTACATCAGATTAACAAGTGGCACTGACAGGACACACTC 204
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1328 GCCGCAATGTTCACATTTGATCCGAATTCATCAGATGACCTTTCAAGACACTGTG 1387
QY 205 GCCAAAGATTACTGCAGCAGCTTCAGAGAGGAGCTTCATCTGC 249
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1388 ACCGTCAAGATTTCAGTAGAAGATGCCAGTACGCTCCCATGTTTC 1432

```

RESULT 4  
 V02032/c  
 ID V02032 standard; DNA; 40352 BP.  
 AC V02032;  
 DT 12-JUN-1998 (first entry)  
 DE MAGE-B cluster DNA sequence.  
 DE MAGE-B cluster DNA; testicular seminoma; cancer; screening; leukaemia;  
 KW carcinoma; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT prim\_transcript 3266..7979  
 FT /\*tag= a  
 FT exon 3266..3364  
 FT /\*tag= b  
 FT /\*note= "exon 1 of B2 gene"  
 FT intron 3365..6277  
 FT /\*tag= c  
 FT /\*note= "intron 1 of B2 gene"  
 FT exon 6278..7979  
 FT /\*tag= d  
 FT /\*note= "exon 2 of B2 gene"  
 FT CDS 6283..7224  
 FT /\*tag= e  
 FT /\*note= "coding sequence for B2 gene"  
 FT polyA\_signal 7961..7966  
 FT /\*tag= f  
 FT exon 23546..25194  
 FT /\*tag= g  
 FT /\*note= "poly-A signal for B2 gene"  
 FT CDS 23607..24647  
 FT /\*tag= h  
 FT /\*note= "coding sequence for B3 gene"  
 FT polyA\_signal 25152..25157  
 FT /\*tag= i  
 FT /\*note= "poly-A signal for B3 gene"

```

FT      prim_transcript 29748..31827
FT      /tag-1
FT      /label-B4 gene
FT      CDS
FT      29808..30848
FT      /tag-k
FT      /note-"coding sequence for B4 gene"
FT      polyA_signal
FT      31822..31827
FT      /tag-1
FT      /note-"poly-A signal for B4 gene"
FT      prim_transcript 31403..39691
FT      /tag-m
FT      /label-B1 gene
FT      exon
FT      31403..31474
FT      /tag-n
FT      /note-"exon 1 of B1 gene"
FT      intron
FT      31475..33957
FT      /tag-o
FT      /note-"intron 1 of B1 gene"
FT      exon
FT      33958..34062
FT      /tag-p
FT      /note-"exon 2 of B1 gene"
FT      intron
FT      34063..35056
FT      /tag-q
FT      /note-"intron 2 of B1 gene"
FT      exon
FT      35057..35139
FT      /tag-r
FT      /note-"exon 3 of B1 gene"
FT      intron
FT      35140..38087
FT      /tag-s
FT      /note-"intron 3 of B1 gene"
FT      exon
FT      38088..39691
FT      /tag-t
FT      /note-"exon 4 of B1 gene"
FT      CDS
FT      38148..39191
FT      /tag-u
FT      /note-"coding sequence for B1 gene"
FT      polyA_signal
FT      39674..39679
FT      /tag-v
FT      /note-"poly-A signal for B1 gene"
FT      WO9746710-A1.
FT      PD 11-DEC-1997.
FT      PR 05-JUN-1997: US-09774.
FT      PR 25-APR-1997: US-846111.
FT      PR 05-JUN-1996: US-658578.
FT      PA (LUDW-) LUDWIG INST CANCER RES.
FT      PA Boon T, Brasseur F, Lurquin C;
FT      DR MPI: 98-042216/04.
FT      PT Screening for the possibility of testicular seminoma and other
FT      cancers - using MAGE-Xp gene primers
FT      PS Claim 7, Pages 22-34; 40pp; English.
FT      CC This is the MAGE-B cluster DNA sequence. The invention provides a novel
FT      CC method for screening for the possibility of a testicular seminoma,
FT      CC non-small cell lung carcinoma, melanoma, breast cancer, sarcoma, or
FT      CC leukaemia in a sample. The method comprises contacting the sample with
FT      CC at least one nucleic acid molecule which hybridises to mRNA corresponding
FT      CC to an MAGE-Xp gene, and determining hybridisation as a determination of
FT      CC possible presence of testicular seminoma, non-small cell lung carcinoma
FT      CC melanoma, breast cancer, sarcoma, or leukaemia in the sample. By assaying
FT      CC for the MAGE-B1 gene, or MAGE-B2 gene, both contained within this 40352
FT      CC base pair MAGE-B cluster DNA sequence, the presence of such cancers in a
FT      CC sample can be determined. The genomic DNA that encodes the MAGE-B2 gene
FT      CC consists of nucleotides 3266-7979 of this MAGE-B1 cluster DNA sequence.
FT      CC The nucleic acid molecule that encodes a MAGE-B1 gene, consists of, in
FT      CC 5' to 3' order, nucleotides 31403-31474, 33958-34062, 35057-35139, and
FT      CC 38088-39691, nucleotides 31403-31474, 33958-34062, and 38088-39691,
FT      CC nucleotides 35057-35139 and 38088-39691, and nucleotides 33958-34062 and
FT      CC 38088-39691 of this sequence. The primers specific for the MAGE-B
FT      CC gene can be used in a kit to amplify a MAGE-B gene.
FT      SQ Sequence 40352 BP; 11599 A; 8654 C; 9772 G; 10327 T;

```

Query Match 11.2%; Score 28.2; DB 1; Length 40352;  
Best Local Similarity 53.1%; Pred. No. 21;

```

Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 73 CAAGCAGCCAGCTTTCTTCCTGCTGCAGACAGTGTAGCTGGGCTCTGGGTGGATGCTG 132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1340 CAACATGACCACTTTTCTTACTGCAAACTTCAGAGGAGGAGCCGTGTCCGGAGTG 1281
QY 133 ACAATTGACAGTCTGCTCCCATGTACAAAGTGAAGTCACTATACAGTAAACATGG 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1280 ATGTGAGTATGAGGAGCACTGAGCAAGGCAAGAGCTCCAGCAGAGTACTGTG 1228

RESULT 5
Q77789
ID Q77789 standard; DNA; 5211 BP.
AC Q77789;
DE 26-JUN-1995 (first entry)
DE Pre-pro-cobra C3 coding sequence.
KW Cobra; C3; third component of complement; human; mouse; rat;
KW X.laavis; pre-pro molecule; beta chain; alpha chain; codon usage;
KW G+C content; Immune response; host defence; ss.
OS Naja naja.
FH Key Location/Qualifiers
FT cds 9..4964
FT /tag-a
FT /product- Pre-pro-cobra C3
FT signal_peptide 9..74
FT /tag-b
FT mat_peptide 75..4961
FT /tag-c
FT misc_difference 480..482
FT /tag-d
FT /codon-seq:GAA, aa:Asp
FT misc_difference 483..485
FT /tag-e
FT /codon-seq:CAA, aa:Lys
PN WO9423024-A.
PD 13-OCT-1994.
PE 07-APR-1994: U03441.
PR 07-APR-1993: US-043747.
PR (GEOU ) UNIV GEORGETOWN.
PI Bredehorst R, Fritzingler DC, Vogel C;
DR MPI: 94-333186/41.
DR P-PSDB: R63222.
PT DNA encoding cobra C3, CVF 1 and CVF 2 - which are used in the
PT treatment of cancer
PT Claim 1; Fig 2A-2L; 155pp; English.
PS This sequence encodes the cobra C3 (third component of complement).
CC The cDNA sequence of cobra C3 shows a high sequence homology with C3
CC molecules from human, mouse, rat and X. laevis. Cobra C3 is
CC synthesised as a pre-pro molecule that is subsequently processed
CC into the mature two-chain protein by removing the signal peptide and
CC the four Arg residues between the beta and alpha chain. The alpha
CC chain comprises 992 amino acids and the beta chain comprises 633
CC residues, being 12 residues shorter than the human beta chain. Cobra
CC C3 has a different codon usage compared to mammalian C3 mRNAs. The
CC G+C content of all known mammalian C3 mRNAs is more than 53%. The
CC G+C content of cobra C3 mRNA is significantly lower at 43%. The
CC significance of this difference is not known. C3 is thought to have
CC important functions in the immune response and host defence.
SQ Sequence 5211 BP; 1612 A; 1042 C; 1201 G; 1356 T;

Query Match 11.0%; Score 27.8; DB 1; Length 5211;
Best Local Similarity 53.2%; Pred. No. 12;
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 102 AGTTGAGCTGGGGTCTCTGGGTGGATGTGATTTGACAGTGTCCCATTAAG 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1730 AGGGGATCCAGGCTGGGGTGGTCTGTGTGTGACAAAGCATATATGTTCTCA 1789
QY 162 TGAATCATACAGATTAACAGTGCAGTGCAGAGGACAGTGCACCAAG 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1790 TGATTAATATAGATTAGCCAAAGCTAAGATATGGGACACAAATGAAAAGG 1840

```

```
RESULT 6
ID 077790 standard; DNA: 5924 BP.
AC 077790;
DE 30-JUN-1995 (first entry)
DE Cobra CVF1 coding sequence.
KW Cobra; C3; third component of complement; human; mouse; rat;
KW X. laevis; pre-pro molecule; beta chain; alpha chain; codon usage;
KW G+C content; immune response; host defence; tumour; ss.
OS Naja naja.
FH Key
FT Location/Qualifiers
FT cds
FT 4..4932
FT /tag= a
FT /product= CVF1 fragment
FT signal_peptide
FT 4..69
FT /tag= b
FT mat_peptide
FT 70..4929
FT /tag= c
PN MO9423024-A.
PD 13-OCT-1994.
PR 07-APR-1994; U03441.
PR 07-APR-1993; US-043747.
PI (GEOU) UNIV GEORGETOWN.
PI Bredehorst R, Fritzing DC, Vogel C;
PI WPI: 94-333186/41.
DR P-PSDB: R63223.
PT DNA encoding cobra C3, CVF 1 and CVF 2 - which are used in the
PT treatment of cancer
PS Claim 3; Fig 2M-2J(10); 155pp; English.
CC This sequence encodes the cobra venom factor 1 (CVF1). The CVF1
CC protein is contained within cobra venom and is a structural analog of
CC cobra C3 (third component of complement). It is thought that CVF1 and
CC C3 are derived from different but related genes. The thioester site
CC and factor H binding site of C3 and CVF1 are very similar, even though
CC neither are actually present in the mature CVF1 protein. Factors H and
CC I are able to regulate the activity of complement by dissociating C3b,
CC Bb, and by cleaving C3b. CVF1 is resistant to this regulation. CVF1
CC and CVF2 may be used in the treatment of cancer by linking them to
CC an antibody which recognises a specific tumour cell. As CVF1 and CVF2
CC are insensitive to factor H control this method will lead to the
CC selective destruction of cancer cells.
CC Sequence 5924 BP; 1817 A; 1196 C; 1316 G; 1595 T;
SQ

Query Match 11.0%; Score 27.8; DB 1; Length 5924;
Best Local Similarity 53.2%; Pred. No. 13;
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 102 AGTTGACCTGGGTCCTGGGTTGGGATGTGACATTGACAGTGTGCCATGTACAAG 161
DB 1707 AGGGGATCCAGTGTCTGGTGTGGTCTGTGCTGTGACAAAGCAGATATGTTCTCAA 1766
QY 162 TGAACCTACATACATAAACAAGTGCAGTGACAGGAGCACAACCTCGCCAAAAG 212
DB 1767 TGATTAATAATATAGATTAGCCAAAGTATGAGACACACATAGAAAAGAG 1817

RESULT 7
ID 077791
ID 077791 standard; DNA: 4138 BP.
AC 077791;
DE 05-JUL-1995 (first entry)
DE Cobra partial CVF2 coding sequence.
KW Cobra; C3; third component of complement; human; mouse; rat;
KW X. laevis; pre-pro molecule; beta chain; alpha chain; codon usage;
KW immune response; host defence; tumour; ss.
OS Naja naja.
FH Key
FT Location/Qualifiers
FT cds
FT 3..4004
FT /tag= a
PN MO9423024-A.
```

```
PD 13-OCT-1994.
PR 07-APR-1994; U03441.
PR 07-APR-1993; US-043747.
PI (GEOU) UNIV GEORGETOWN.
PI Bredehorst R, Fritzing DC, Vogel C;
PI WPI: 94-333186/41.
DR P-PSDB: R63224.
PT DNA encoding cobra C3, CVF 1 and CVF 2 - which are used in the
PT treatment of cancer
PS Claim 5; Fig 2M-2V; 155pp; English.
CC This sequence encodes the C-terminal portion of cobra venom factor
CC 2 (CVF2). A full length sequence encoding pre-pro CVF2 may be
CC constructed by ligating this sequence to the 3' end of the 1-964 bp
CC fragment derived from CVF1 (see also 077790). CVF2 is active in both
CC its glycosylated and unglycosylated states. Factors H and I are able
CC to regulate the activity of complement by dissociating C3b, Bb, and by
CC cleaving C3b. CVF2 is resistant to this regulation. CVF1 and CVF2 may
CC be used in the treatment of cancer by linking them to an antibody which
CC recognises a specific tumour cell. As CVF1 and CVF2 are insensitive to
CC factor H control this method will lead to the selective destruction of
CC cancer cells.
SQ Sequence 4138 BP; 1303 A; 823 C; 944 G; 1068 T;

Query Match 11.0%; Score 27.8; DB 1; Length 4138;
Best Local Similarity 47.2%; Pred. No. 11;
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
```

```
QY 102 AGTTGACCTGGGTCCTGGGTTGGGATGTGACATTGACAGTGTGCCATGTACAAG 161
DB 755 AGGGGATCCAGTGTCTGGTGTGGTCTGTGCTGTGACAAAGCAGATATGTTCTCAA 814
QY 162 TGAACCTACATACATAAACAAGTGCAGTGACAGGAGCACAACCTCGCCAAAAG 212
DB 815 TGATTAATAATATAGATTAGCCAAAGTATGAGACACACATAGAAAAGAG 865

RESULT 8
ID V66832/c
ID V66832 standard; DNA: 2990 BP.
AC V66832;
DE 05-JAN-1999 (first entry)
DE Zea mays soluble starch synthase gene SSS1052 and SSS64.
KW Zea mays; US yellow-dent corn line; maize; soluble starch synthase;
KW SSS; glycogen biosynthetic pathway; branching enzyme; ss.
OS Zea mays.
PN US5824790-A.
PD 20-OCT-1998.
PR 15-DEC-1995; 572951.
PR 15-DEC-1995; US-572951.
PR 21-JUN-1994; US-263921.
PR 29-NOV-1994; US-346602.
PI (ZENE) ZENECA LTD.
PI Guan H, Keeling PL, Knight ME;
PI WPI: 98-582626/49.
PT Isolated nucleic acid molecule, used to produce transgenic plants -
PT comprises nucleotide sequence encoding polypeptide having soluble
PT starch synthase activity, where polypeptide is encoded by maize gene
PS Claim 1; Column 25-28; 29pp; English.
CC The present sequence represents an isolated nucleic acid molecule which
CC has been isolated and comprises a nucleotide sequence encoding a
CC polypeptide having soluble starch synthase (SSS) activity, where the
CC polypeptide is encoded by a maize gene. The isolated nucleic acid
CC molecule can be used to produce transgenic plants with altered starch
CC production. The transgenic plants produced using the nucleic acid
CC molecule have an enhanced ability to produce structurally-altered starch.
SQ Sequence 2990 BP; 758 A; 655 C; 801 G; 776 T;
```



DB 317 TGTACTCCCTTCAACATGACGCTAGCCCGAGAGATGAAGACGTTACCA 258  
QY 108 GCTGGGGTCTGGGTGGATGATGACATTTGACAGTGTGCGCATGTACAAG 161  
DB 257 CTGCGAGTCTGCGCAAGAAAGTCTTCTTTCACAGAGAAAGCCCATGGACATG 204

RESULT 11  
ID 050187 standard; DNA: 1130 BP.  
AC 050187;  
DT 05-MAY-1994 (first entry)  
KW Porcine factor VIII A2 domain coding sequence.  
KW Blood clotting; coagulation; hybrid human/porcine factor VIII;  
KW anti-hemophilic factor; factor 8; ss.  
OS Sus scrofa.  
FH Key Location/Qualifiers  
FT cds 2. 1093  
FT /tag- a  
FT /product- A2\_domain  
FT /note- open reading frame continues 3' of  
FT nucleotide 1093\*

PN M09320093-A.  
PD 14-OCT-1993.  
PE 07-APR-1993; U03275.  
PR 07-APR-1992; US-864004.  
PA (UTEM-) UNIV EMORY.  
PI Lolilar JS, Runge MS;  
DR P-PSDB: R43259.  
PT Hybrid human and porcine factor VIII - having high coagulant  
PT activity, used for treating patients with factor VIII deficiency  
PT Partic. haemophilia  
PS Claim 29; Page 52-53; 94pp; English.  
CC Hybrid human/porcine factor VIII molecules are claimed. These can be  
CC obtained by (1) substituting a porcine subunit (i.e. heavy or light  
CC chain) for the corresp. human subunit; (2) substituting a porcine  
CC domain (i.e. A1, A2, A3, B, C1 and C2) for the corresp. human domain;  
CC (3) substituting part of a porcine domain for the corresp. human  
CC domain or (4) changing residues in the human sequence to the corresp.  
CC porcine residues. The hybrid factor VIII molecules have greater  
CC clotting activity than human factor VIII and are more resistant to  
CC factor VIII inhibitors. Nucleic acid coding for the A2 domain of  
CC porcine factor VIII, homologous to residues 372-740 in mature human  
CC factor VIII, is also claimed.  
SQ Sequence 1130 BP; 321 A; 271 C; 258 G; 280 T;

Query Match 10.9%; Score 27.4; DB 1; Length 1130;  
Best Local Similarity 57.6%; Pred. No. 8.6;  
Matches 49; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 125 GGATGGTGCATTTGACAGTCTGCCCATGTACAAGTGAATCATACAGATAAAGCTG 184  
DB 353 GGTTGGAAACATTTTGAAGACATGTCGACAGAGAGACTTCAAGTATAATG 412

QY 185 GCAGTGCAGGAGACACACTGCCAA 209  
DB 413 ACACTGACTGTGGAAGATGGCCCA 437

RESULT 12  
ID 066616 standard; CDNA: 1623 BP.  
AC 066616;  
DT 10-NOV-1994 (first entry)  
DE Sequence of porcine factor VIII.  
KW Factor VIII; haemostasis; haemophilia A; clotting cascade;  
KW fibrinogen; fibrin; thrombin; proteolytic enzyme; co-factor; ss.  
OS Sus scrofa.  
FH Key Location/Qualifiers  
FT cds 1. 1623  
FT /tag- a

PN M09411503-A.  
PD 26-MAY-1994.  
PE 01-OCT-1993; U09438.  
PR 13-NOV-1992; US-976086.  
PR 14-SEP-1993; US-121202.  
PA (GENEX) GENETICS INST INC.  
PI Kaufman RJ, Piltman D, Rehemtulla A, Wozney JM;  
DR WPI: 94-183504/22.  
DR P-PSDB: R55353.  
PT Nucleic acid encoding porcine factor VIII - used to obtain  
PT porcine and human-porcine chimeric factor VIII for treating  
PT haemophilia  
PS Claim 2; Page 43-45; 61pp; English.  
CC A partial genomic clone for the porcine factor VIII B domain was  
CC obtd. as described in US Patent No. 4,757,006, and included an  
CC intron connected to another exon presumed to be in the A2 domain.  
CC An oligo probe corresp. to the thrombin cleavage site at the COOH  
CC terminus of porcine factor VIII A2 domain was designed corresp. to  
CC 066617. This sequence (P831) was derived from a porcine genomic  
CC clone that contains the protein sequence EPRSL and corresp. to  
CC AAs 738-743 of the human protein. This probe, in combination with  
CC degenerate oligo probes made against various sequences of the A1  
CC domain of human factor VIII, was used to amplify the corresp.  
CC sequences for porcine factor VIII. Examples of sequences used to  
CC isolate a porcine fragment are P857 (066618) and P854 (066619).  
CC RNA was prepd. from porcine liver and converted to cDNA. It was  
CC PCR amplified with 5' oligo P854 and 3' oligo P831.  
CC The results are set forth in 066616. Chimeric forms of factor  
CC VIII include those where various domains of the human factor  
CC factor VIII domains, and include, chimeric forms where the A1  
CC and/or A2 domains, in whole or in part, of the human factor VIII  
CC sequence have been replaced, in whole or in part, by the A1 and/or  
CC the A2 domains of porcine factor VIII. Specifically provided are  
CC chimeric factor VIII sequences comprising the A1, A2, A3, B, C1 and  
CC C2 human domains as set forth in 066615, where the A1 and/or A2  
CC domains, as well as other segments, such as the regions corresp. to the  
CC AA numbers 336-372, 336-740, 372-740, 700-740 and combinations of these  
CC regions have been replaced in whole or in part with porcine factor  
CC VIII sequences as set forth in 066616 and R55353.  
SQ Sequence 1623 BP; 434 A; 403 C; 389 G; 397 T;

Query Match 10.9%; Score 27.4; DB 1; Length 1623;  
Best Local Similarity 57.6%; Pred. No. 10;  
Matches 49; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 125 GGATGGTGCATTTGACAGTCTGCCCATGTACAAGTGAATCATACAGATAAAGCTG 184  
DB 883 GGTGGAAACATTTTGAAGACATGTCGACAGAGAGACTTCAAGTATAATG 942

QY 185 GCAGTGCAGGAGACACACTGCCAA 209  
DB 943 ACACTGACTGTGGAAGATGGCCCA 967

RESULT 13  
ID V12114 standard; CDNA to mRNA; 1130 BP.  
AC V12114;  
DT 17-JUN-1998 (first entry)  
DE Sus scrofa factor VIII A2 domain coding region.  
KW factor VIII; recombinant; modified; haemophilia; treatment; ds.  
OS Sus scrofa.  
FH Key Location/Qualifiers  
FT cds 1. 1130  
FT /tag- a  
FT /product- factor VIII fragment including A2 domain  
FT /note- "no stop codon given"

PN M09749725-A1.  
PD 31-DEC-1997.  
PE 26-JUN-1997; U11155.  
PR 26-JUN-1996; US-670707.

PA (UYEM-) UNIV EMORY.  
 PI Lollar JS;  
 DR WPI: 98-077108/07.  
 PR P-PSDB: W44134.  
 PT New modified factor VIII molecules - having reducing immunogenicity  
 PS Example: Page 71: 126pp: English.  
 CC The sequence is that encoding a fragment of porcine factor VIII  
 CC including the entire A2 domain. This can be used in the  
 CC production of modified factor VIII. The factor VIII molecules  
 CC have coagulant activity and can be used for treating factor VIII  
 CC deficiency, particularly for treating patients with haemophilia.  
 CC The products can also be used in detection and diagnosis. This  
 CC modified factor VIII has less immunoreactivity with naturally  
 CC occurring inhibitory antibodies to factor VIII and may be less apt  
 CC to elicit the production of antibodies to factor VIII than human  
 CC factor VIII. Some of the hybrid factor VIII molecules have specific  
 CC activity greater than that of human factor VIII and equal to or greater  
 CC than that of porcine factor VIII.  
 SQ Sequence 1130 BP; 321 A; 271 C; 258 G; 280 T;

Query Match 10.9%; Score 27.4; DB 1; Length 1130;  
 Best Local Similarity 57.6%; Pred. No. 8.6;  
 Matches 49; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
 QY 125 GGATGGTACATTGACAGTCTGCCCATGTACAAAGTGAAGTACATACAGATAACAGTG 184  
 DB 353 GGTGGAAACATTGTGAAGACATGCCAATTCGCCAGAGAGACTTCAAGTATAAATGG 412  
 QY 185 GCAGTGACAGGACACACTCGCCAA 209  
 DB 413 ACAGTGACTGTGGAAGATGGCCAA 437

RESULT 14  
 ID V12113 standard; cDNA to mRNA; 6402 BP.  
 AC V12113;  
 DT 17-JUN-1998 (first entry)  
 DE Sus scrofa factor VIII coding region.  
 KM factor VIII; recombinant; modified; haemophilia; treatment; ds.  
 OS Sus scrofa.  
 FH Key Location/Qualifiers  
 FT CDS 1..6402  
 FT /\*tag= a  
 FT /product= factor VIII  
 PN MO9749725-A1.  
 PD 31-DEC-1997.  
 PF 26-JUN-1997; U11155.  
 PR 26-JUN-1996; US-670707.  
 PA (UYEM-) UNIV EMORY.  
 PI Lollar JS;  
 DR WPI: 98-077108/07.  
 PR P-PSDB: W44133.  
 PT New modified factor VIII molecules - having reducing immunogenicity  
 PS Claim 24; Pages 94-102; 126pp: English.  
 CC The sequence is that encoding porcine factor VIII. It can be used in  
 CC the production of modified factor VIII. The factor VIII molecules  
 CC have coagulant activity and can be used for treating factor VIII  
 CC deficiency, particularly for treating patients with haemophilia.  
 CC The products can also be used in detection and diagnosis. This  
 CC modified factor VIII has less immunoreactivity with naturally  
 CC occurring inhibitory antibodies to factor VIII and may be less apt  
 CC to elicit the production of antibodies to factor VIII than human  
 CC factor VIII. Some of the hybrid factor VIII molecules have specific  
 CC activity greater than that of human factor VIII and equal to or greater  
 CC than that of porcine factor VIII.  
 SQ Sequence 6402 BP; 1739 A; 1618 C; 1614 G; 1431 T;

Query Match 10.9%; Score 27.4; DB 1; Length 6402;  
 Best Local Similarity 57.6%; Pred. No. 18;  
 Matches 49; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 125 GGATGGTACATTGACAGTCTGCCCATGTACAAAGTGAAGTACATACAGATAACAGTG 184  
 DB 1537 GGTGGAAACATTGTGAAGACATGCCAATTCGCCAGAGAGACTTCAAGTATAAATGG 1596  
 QY 185 GCAGTGACAGGACACACTCGCCAA 209  
 DB 1597 ACAGTGACTGTGGAAGATGGCCAA 1621

RESULT 15  
 ID V12142 standard; cDNA to mRNA; 4334 BP.  
 AC V12142;  
 DT 17-JUN-1998 (first entry)  
 DE Homo sapiens factor VIII coding region lacking B domain.  
 KM factor VIII; recombinant; modified; haemophilia; treatment; ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 3..4334  
 FT /\*tag= a  
 FT /product= factor VIII lacking B domain  
 PN MO9749725-A1.  
 PD 31-DEC-1997.  
 PF 26-JUN-1997; U11155.  
 PR 26-JUN-1996; US-670707.  
 PA (UYEM-) UNIV EMORY.  
 PI Lollar JS;  
 DR WPI: 98-077108/07.  
 PR P-PSDB: W44132.  
 PT New modified factor VIII molecules - having reducing immunogenicity  
 PS Example: Pages 109-114; 126pp: English.  
 CC The sequence is that encoding modified human factor VIII which lacks  
 CC the B domain. Such modified factor VIII molecules have  
 CC coagulant activity and can be used for treating factor VIII  
 CC deficiency, particularly for treating patients with haemophilia.  
 CC The products can also be used in detection and diagnosis. This  
 CC modified factor VIII has less immunoreactivity with naturally  
 CC occurring inhibitory antibodies to factor VIII and may be less apt  
 CC to elicit the production of antibodies to factor VIII than human  
 CC factor VIII. Some of the hybrid factor VIII molecules have specific  
 CC activity greater than that of human factor VIII and equal to or greater  
 CC than that of porcine factor VIII.  
 SQ Sequence 4334 BP; 1109 A; 1126 C; 1085 G; 1014 T;

Query Match 10.9%; Score 27.4; DB 1; Length 4334;  
 Best Local Similarity 57.6%; Pred. No. 15;  
 Matches 49; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
 QY 125 GGATGGTACATTGACAGTCTGCCCATGTACAAAGTGAAGTACATACAGATAACAGTG 184  
 DB 1539 GGTGGAAACATTGTGAAGACATGCCAATTCGCCAGAGAGACTTCAAGTATAAATGG 1598  
 QY 185 GCAGTGACAGGACACACTCGCCAA 209  
 DB 1599 ACAGTGACTGTGGAAGATGGCCAA 1623

Search completed: August 6, 1999, 04:02:41  
 Job time: 1513 sec

---

This Page Blank (uspto)



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:34:36 ; Search time 3268.17 Seconds

(without alignments)  
245.225 Million cell updates/sec

Title: US-09-049-696-6

Perfect score: 252  
Sequence: 1 CAAAGATGTGTGTGTAGT.....GAGGAGCTGCATCTGCAGC 252

Scoring table: OLIGO\_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl:.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pl3.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_sc.\*  
14: gb\_sts.\*  
15: gb\_sy.\*  
16: gb\_un.\*  
17: gb\_vl.\*  
18: em\_fun.\*  
19: em\_hcg.\*  
20: em\_hum1.\*  
21: em\_hum2.\*  
22: em\_in.\*  
23: em\_om.\*  
24: em\_or.\*  
25: em\_ov.\*  
26: em\_pat.\*  
27: em\_ph.\*  
28: em\_pl.\*  
29: em\_ro.\*  
30: em\_sts.\*  
31: em\_sy.\*  
32: em\_un.\*  
33: em\_vl.\*  
34: gb\_hcg1.\*  
35: gb\_hcg2.\*  
36: gb\_in1.\*  
37: gb\_in2.\*  
38: em\_ba1.\*  
39: em\_ba2.\*  
40: em\_hum3.\*  
41: em\_hum4.\*  
42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	252	100.0	3311	11	AF039400	Homo sapi

2	252	100.0	2826	42	AF127036	AF127036 Homo sapi
3	208	82.5	35278	11	AF039401	AF039401 Homo sapi
4	19	7.5	4914	2	AF001171	AF001171 Ralstonia
5	19	7.5	279	14	G47065	G47065 Z15539_1 Ze
6	18	7.1	24813	4	FRICRA1	AF110525 Fugu rubr
7	18	7.1	13801	4	FRICRA2	AF110526 Fugu rubr
8	18	7.1	94319	9	HS411T10	Z97352 Human DNA S
9	18	7.1	148416	10	HSAC002089	AC002089 Human BAC
10	18	7.1	174952	11	AC002460	AC002460 Human BAC
11	18	7.1	122816	11	AC002523	AC002523 Homo sapi
12	18	7.1	192548	11	AC005296	AC005296 Homo sapi
13	18	7.1	107190	11	AC005611	AC005611 Homo sapi
14	18	7.1	19839	11	AF094481	AF094481 Homo sapi
15	18	7.1	2534	12	AB011156	AB011156 Mus muscu
16	18	7.1	161111	34	HSJ117516	MS8164 Mesocricetu
17	18	7.1	120479	35	AC006945	AL049538 Homo sapi
18	18	7.1	120479	35	AC006945	AC006945 Mus muscu
19	17	6.7	486	1	CAC013353	AJ012353 Clostridi
20	17	6.7	86896	1	RCU57682	U57682 Rhodobacter
21	17	6.7	5071	3	SSSDCBR	X64315 S.scrofa MR
22	17	6.7	3255	4	AB002404	AB002404 Gallus ga
23	17	6.7	5665	4	AB01870783	AB018709 Gallus ga
24	17	6.7	13000	4	AF058764	AF058764 Xenopus 1
25	17	6.7	8000	4	XLAJ5076	AJ005076 Xenopus 1
26	17	6.7	4530	5	A36408	A36408 Sequence 1
27	17	6.7	5420	9	AB023150	AB023150 Homo sapi
28	17	6.7	100000	9	AP000073	AP000073 Homo sapi
29	17	6.7	198161	9	HS232K4	AL021938 Homo sapi
30	17	6.7	149140	9	HS256G22	AL022097 Homo sapi
31	17	6.7	161076	9	HS272T12	Z82194 Human DNA s
32	17	6.7	110508	9	HS283E3	AL031282 Human DNA
33	17	6.7	14776	9	HDMCPIIE	J02843 Human cytoc
34	17	6.7	7288	10	HSAN237839	AJ237839 Homo sapi
35	17	6.7	44057	10	HSCN37F10	269714 Human DNA s
36	17	6.7	759	10	HSFPGS4	U40866 Human foilyl
37	17	6.7	16302	10	SOE010154	AJ010154 Saguinus
38	17	6.7	167101	11	AC002094	AC002094 Genomic s
39	17	6.7	161326	11	AC004067	AC004067 Homo sapi
40	17	6.7	127514	11	AC004417	AC004417 Homo sapi
41	17	6.7	190434	11	AC005771	AC005771 Homo sapi
42	17	6.7	37225	11	AC005954	AC005954 Homo sapi
43	17	6.7	44419	11	AC006506	AC006506 Homo sapi
44	17	6.7	162691	12	AC003061	AC003061 Mouse Chr
45	17	6.7	274433	35	AC007368	AC007368 Homo sapi

## ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	TITLE	REFERENCE	AUTHORS	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE
1	AF039400	AF039400	3311 bp	3311	AF039400.1	GI:4009457	complete cds.	Homo sapiens	Human.	Genomic cloning, molecular characterization, and functional analysis of human CACAL, the first human member of the family of Ca2+-activated Cl- channel proteins	Gruber A.D., Elble, R.C., Ji, H.L., Schreuer, K.D., Fuller, C.M. and Pauli, B.U.	Genomics 54 (2), 200-214 (1998)	99047526	2 (bases 1 to 3311)	Gruber A.D., Elble, R. and Pauli, B.U.	Direct Submission

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES Location/Qualifiers

source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="pid:94009458"  
/db\_xref="gi:4009458"  
/translation="MGPRKSSVFLLILHLEGALSNSLIQLNNNGYEGIVAIIDPNV  
EDETLIQIKDMVTQASLYLEFATGRKRYFNKVAIILPEWTAKADYVRFLETKNA  
DVLVAESTPNDPEPTEDMNGCGEGERIHLIPDIAGKLAIEYGPQRAFEVEMAH  
LRMGVDEYNDEKFIYLSNGRIQAVRCSAGITGLNVKRCOGSCYTRCTFNKVTGL  
YKGECEFLVLOSROTERKASIMFAQHVDSIERFCEQNHNRKAPKONKCNLRSTWEVI  
RDSDEFKTTPTTOPPNTFSLQIGORIVCYLKRSGSMATGNRLNRNOGOLF  
LQVLELGSWGMVTFDSAAHVOSLIIQINSQSDRDLAKRPAASGCTSCGLRSA  
FTVIRKRYPTDGSSEIYLTDGEDNTISGCEFNKOSGAIHTHYALGPSAQDELELSK  
MTGGLQIYASDVQVQNGNLIDAFGALSNGNVAVSQISLESKGLTLQNSQMNQTVIY  
DVSQKDTLELITWTTPQPIILMDPSGQKOGFVVDKNTKMAVLIQIPGAKGVTVK  
SLQASSQTLILVITSRSANATLPITVTSKTNKIDSKPSPLVYANIRQASPIILRA  
SVTALIESVNGKTVLELLDNGAGADATKDDGVSRFTYDNGRSYKVRALGGVN  
AARAVIPIQSGALYIPGMIENDEIOMNPPEIKNDDVOHKOVCSRTSSGGSPVNS  
DVNAPILPDLFPPGQITDLKAEIHGGSILNTMTAGDDYDHGTAKYIIRISITSLD  
LRKNESLQVNTTALIRKANESEVFLFRPENITTFENGITLFIQLQAVDKVLKSEI  
SNIAKVSLEIPQTPPETPSPDETSAPCPNHIINSTIPGIIHLIKMKWIGELQLSIA

BASE COUNT 1028 a 692 c 742 g 849 t

ORIGIN

Query Match 100.0%; Score 252; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 6,7e-138;  
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGAATGTTGGTGTAGTCTTGACAAATCTGGAAGCATGGGACATGCTAACCGCCTC 60  
DB 1261 CAAAGAAATTGTGTCTTATCTTGCAAAATCTGGAAGCATGGGACATGCTAACCGCCTC 1320  
QY 61 AATCGACTGAATCAAGCAGCGCAGCTTTTCTCTGTCGACAGCAGTTGAGCTGGGCTCTGG 120  
DB 1321 AATCGACTGAATCAAGCAGCGCAGCTTTTCTCTGTCGACAGCAGTTGAGCTGGGCTCTGG 1380  
QY 121 GTTGGGATGGTACATTTGACAGTGTCTGCCCATGTACAAAGTGAACATACAGATAAC 180  
DB 1381 GTTGGGATGGTACATTTGACAGTGTCTGCCCATGTACAAAGTGAACATACAGATAAC 1440  
QY 181 AGTGCAGTGAAGCAGGACACACTCGCCAAAAGATTACCTGACGACGCTTCACAGAGGAGC 240  
DB 1441 AGTGCAGTGAAGCAGGACACACTCGCCAAAAGATTACCTGACGACGCTTCACAGAGGAGC 1500  
QY 241 TCCATCTGCAGC 252  
DB 1501 TCCATCTGCAGC 1512

RESULT 2  
AF127036 2826 bp mRNA PRI 16-Apr-1999  
LOCUS AF127036 Homo sapiens calcium-activated chloride channel protein 1 (Caccl1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
MID 94585468  
VERSION AF127036-1 GI:4585468

KEYWORDS  
SOURCE

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 2826)  
Agnel,M., Verma,T. and Culouscou,J.-M.  
Cloning of three human homologs of bovine epithelial chloride  
channel

JOURNAL  
Unpublished

REFERENCE  
2 (bases 1 to 2826)  
Agnel,M. and Culouscou,J.-M.  
Direct Submission

JOURNAL  
Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrieres, Rueil-Malmaison 92500, France

FEATURES

source

gene  
CDS  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
/gene="Caccl1"  
5. .2749  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="A025487.1"  
/db\_xref="pid:94585469"  
/db\_xref="gi:4585469"  
/translation="MGPRKSSVFLLILHLEGALSNSLIQLNNNGYEGIVAIIDPNV  
EDETLIQIKDMVTQASLYLEFATGRKRYFNKVAIILPEWTAKADYVRFLETKNA  
DVLVAESTPNDPEPTEDMNGCGEGERIHLIPDIAGKLAIEYGPQRAFEVEMAH  
LRMGVDEYNDEKFIYLSNGRIQAVRCSAGITGLNVKRCOGSCYTRCTFNKVTGL  
YKGECEFLVLOSROTERKASIMFAQHVDSIERFCEQNHNRKAPKONKCNLRSTWEVI  
RDSDEFKTTPTTOPPNTFSLQIGORIVCYLKRSGSMATGNRLNRNOGOLF  
LQVLELGSWGMVTFDSAAHVOSLIIQINSQSDRDLAKRPAASGCTSCGLRSA  
FTVIRKRYPTDGSSEIYLTDGEDNTISGCEFNKOSGAIHTHYALGPSAQDELELSK  
MTGGLQIYASDVQVQNGNLIDAFGALSNGNVAVSQISLESKGLTLQNSQMNQTVIY  
DVSQKDTLELITWTTPQPIILMDPSGQKOGFVVDKNTKMAVLIQIPGAKGVTVK  
SLQASSQTLILVITSRSANATLPITVTSKTNKIDSKPSPLVYANIRQASPIILRA  
SVTALIESVNGKTVLELLDNGAGADATKDDGVSRFTYDNGRSYKVRALGGVN  
AARAVIPIQSGALYIPGMIENDEIOMNPPEIKNDDVOHKOVCSRTSSGGSPVNS  
DVNAPILPDLFPPGQITDLKAEIHGGSILNTMTAGDDYDHGTAKYIIRISITSLD  
LRKNESLQVNTTALIRKANESEVFLFRPENITTFENGITLFIQLQAVDKVLKSEI  
SNIAKVSLEIPQTPPETPSPDETSAPCPNHIINSTIPGIIHLIKMKWIGELQLSIA

BASE COUNT 875 a 623 c 632 g 696 t

ORIGIN

Query Match 100.0%; Score 252; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 6,5e-138;  
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGAATGTTGGTGTAGTCTTGACAAATCTGGAAGCATGGGACATGCTAACCGCCTC 60  
DB 914 CAAAGAAATTGTGTCTTATCTTGCAAAATCTGGAAGCATGGGACATGCTAACCGCCTC 973  
QY 61 AATCGACTGAATCAAGCAGCGCAGCTTTTCTCTGTCGACAGCAGTTGAGCTGGGCTCTGG 120  
DB 974 AATCGACTGAATCAAGCAGCGCAGCTTTTCTCTGTCGACAGCAGTTGAGCTGGGCTCTGG 1033  
QY 121 GTTGGGATGGTACATTTGACAGTGTCTGCCCATGTACAAAGTGAACATACAGATAAC 180  
DB 1034 GTTGGGATGGTACATTTGACAGTGTCTGCCCATGTACAAAGTGAACATACAGATAAC 1093  
QY 181 AGTGCAGTGAAGCAGGACACACTCGCCAAAAGATTACCTGACGACGCTTCACAGAGGAGC 240  
DB 1094 AGTGCAGTGAAGCAGGACACACTCGCCAAAAGATTACCTGACGACGCTTCACAGAGGAGC 1153  
QY 241 TCCATCTGCAGC 252  
DB 1153 TCCATCTGCAGC 1153

DB 1154 TCATCTGCAGC 1165

RESULT 3  
LOCUS AF039401 35278 bp DNA PRI 15-DEC-1998  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene, complete cds.

ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schneur,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

FEATURES  
source  
1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="11 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=-1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=-2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="pid:g4009460"  
/db\_xref="gi:4009460"  
/translation="MGPKSSVFIIILHLEGALSNSLIQLNNNGYEGIVAIIDNPVDELIIQIKDWVTOASLYLEFATGKREFKVNALLIETWTKADYVRPKLEYKNA DIVAVSTPPGNDDEPTEDMGNGEGERIHLTPDFIAGKLAEGPOGKAFFHEMAH LRMGVDEYNDEKFTYLSNGRIQAVCSAGITGVVKKCOGSGCYTRKCPNKYVGL YKGCFTVLOSDEKASIMFPAOHVDSIYERCTEDNNHKEAPNRKONCRLASTTEVI RDEDFKTTTPMTTPNPPTSLIIGORIVCLVDKSGSNATGRNLRLNNOGOLF

1QTVPLSGWGMVTFDSAAYQSELIQINSQSDRDTLAKRLPAASGCTSCSGRRA  
FTVIRKKIPTDGESEVILLTDEEDNTISCFENVKOSGAI IHTVALGPSADLEELSK  
MTGGLQTYASDOYONNGLIIDAFGALSNGAVSORSIQLESGLTQSONMNGEIVY  
DSIVGKDLFLITTTTPOPIILMDPSGOKOGFEVDKNTKMAYLQIIGIAKVWTKY  
SLQASSQFLTLVTSRASNAITLPIITVSKTNKPSKPSPLVYANIRQGSPIIRA  
SVTALIESVNGKTVITLLELDNGAGADATKDGAVYSRPTTDTNGRSVVKRAGGVN  
AARRVITQOOSGALYIPGVIINDEIQMNPPEIKNDVQKQVCFSTSSGGSFVAS  
DVNPAPIDPLFPFGQITDLNMEIHGGSLLNLTWAPGDYDHGTHAKTITRISVILD  
LNDKNEISIQVNTALIPKEANSEVEFLKPEENITFEENITDIFAIOADVVDLKEI  
SNIAVSESLFIPQPTPTSPDSEAPCPNIHINSTIPGIIHLKIMMWIGELQISIA  
6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15

BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN

Query Match 82.5%; Score 208; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 7.5e-112;  
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 GACTGTACCGCCCAATGCACTGAATCAAGCAGGCCAGCTTTCCTGCTCAGACGT 104  
DB 19782 GACTGTAAACCCCTCAATGCACTGAATCAAGCAGGCCAGCTTTCCTGCTCAGACGT 19841

QY 105 TAGCTGGGGTCTGGGTGGGATGGATGACATTTGACAGTGGTCCCATGTCAAAAGTGA 164  
DB 19842 TAGCTGGGGTCTGGGTGGGATGGATGACATTTGACAGTGGTCCCATGTCAAAAGTGA 19901

QY 165 ACTCATACAGATAAAGTGGCAGTGACAGGAGACACACTGCCAAAGATTACTGCAGC 224  
DB 19902 ACTCATACAGATAAAGTGGCAGTGACAGGAGACACACTGCCAAAGATTACTGCAGC 19961

QY 225 AGCTCAGGAGGAGGAGTGCATCTGCAGC 252  
DB 19962 AGCTCAGGAGGAGGAGTGCATCTGCAGC 19989

RESULT 4  
AF001171/c DNA BCT 29-JUN-1998  
LOCUS Ralstonia solanacearum protein kinase homolog (pehs) and pehr  
DEFINITION (pehr) genes, complete cds.  
ACCESSION AF001171  
NID 93264837  
VERSION AF001171.1 GI:3264837  
KEYWORDS  
SOURCE Ralstonia solanacearum.  
ORGANISM Ralstonia solanacearum.  
Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
Ralstonia.  
1 (bases 1 to 4914)  
REFERENCE  
AUTHORS Allen, C.  
TITLE Direct Submission  
JOURNAL Submitted (23-APR-1997) Plant Pathology, University of Wisconsin at  
Madison, 1630 Linden Drive, Madison, WI 53705, USA  
COMMENT On Jun 29, 1998 this sequence version replaced gi:2443474.  
FEATURES  
source  
1..4914  
/organism="Ralstonia solanacearum"  
/strain="K60"  
/db\_xref="taxon:305"  
496..2403  
/gene="pehs"  
496..2403  
/function="control polgalacturonase production and  
possibly motility"  
/note="pehs"  
/codon\_start=1  
/transl\_table=1  
/product="protein kinase homolog"  
/protein\_id="AAC24598.1"  
/db\_xref="PID:92443475"  
/db\_xref="GI:2443475"  
/translation="MRLRYAFSRAVALVLLFISIPREHATDLPASVGDAMLST  
LPYLAALILIAAGWWRARFQFVRDLVDLDFLGAYTTLRLSASVAMVLMV  
LAAGLSTLFLTAAYASVVAEPLRMIGDTIDSGLASGLIGLYMAALNM  
YGLSHROYAERLTARELERLQOLVRLAVYMDQGVMLVRADGRVYANPAAAL  
LGVPMAFVSGSVLFDLKGIPHLLETIRQLRQSRHRSRSGTGDADRITD  
LQPTAGRAALRARIRLRLPLSLANRTYLDLSYGATLGLPGEAVGPPRP  
ATETVAGSWADDEVFLNHELDVTLVAVESWERTEDAQOETLASWGRVAVAHQI  
RNPLAISOAAELLDLDGEGEPRLPEGRVETRLRLIRNVRRLDQVAVAHQI  
RPRGRVVOIAQVLPVEYERVARALRAGEATEIHADLVAVDLPGVLPDPAL  
QOVAGNLIDNALRYCRVPSILLAAYPLDTHALVIMNDGPEVSAEQOGLPEPF  
TNDAGTGIGLYMARLCAANDAIIRYGDIALSLDRLTGALTIVAREALPRRFVIT  
LMFOPAVPAR"  
2430..4109  
/gene="pehr"  
2430..4109  
/gene="pehr"  
/function="control polgalacturonase production and  
possibly motility"  
/note="response regulator"  
/codon\_start=1  
/transl\_table=1  
/product="pehr"  
/protein\_id="AAC24599.1"  
/db\_xref="PID:92443476"  
/db\_xref="GI:2443476"  
/translation="MSKAATVREPILVVDDEADIRELLETSLRRMGHDVLAAGLGA  
REALARQFALVLTMDRLDGLGIDLVROLSATDRTPVAVITAYGSAENAVALKGA  
AFDYIAKPLSLDQLSLVNLALGRQDPDPSGADLAERFNLALPGHSAMQERRL  
LRLARSAVIVISGSGKGERARAIHALSARSPREPVNCAIPENIMAEFEFY  
VKGAFTGADSDRQGFQOAHGGTLMDEVADLPTLMQVKKLLRLQDGRVAKIESPD  
PVDRVYCAATHONLARLYAAGRREDI.FYRLNTELMPTLREAEVPIVLAQYLEQ  
LATRGDPFRKRLTRQALQOICATPFPGNRVDVNLLERAVAFRAGESIDVDHGLAG  
SDIERSPLFHRARAHAGHPVPAHLPPVAPGHPAHPGSHGVAHGLVGPDPVGGP  
-DAAAYIIVPGPMGLVPHPAVVPPEPAPAPASPMFVSLPVDLPAVLESVERSVIL

misc\_feature  
4108..4706  
/note="similar to AMPD encoded by Escherichia coli,  
Citrobacter freundii, and Salmonella typhimurium"  
BASE COUNT 651 a 1694 c 1721 g 848 t  
ORIGIN

Query Match 7.5%; Score 19; DB 2; Length 4914;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 AGCCAGCTTTCTGCTG 96  
|||||  
Db 1633 AGCCAGCTTTCTGCTG 1615

RESULT 5  
G47065/c  
LOCUS 279 bp DNA STS 23-MAR-1999  
DEFINITION Z15539.1 zebrafish AB Danio rerio STS genomic clone Z15539 5',  
sequence tagged site.  
ACCESSION G47065  
NID G4493385  
VERSION G47065.1 GI:4493385  
KEYWORDS STS.  
SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinoidae; Cyprinidae; Rasbora; Danio.  
1 (bases 1 to 279)  
Zinif, J., Jackson, D.G. and Fishman, M.C.  
A zebrafish microsatellite map  
Unpublished (1999)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Contact: Mark C. Fishman  
Cardiovascular Research Center  
Massachusetts General Hospital  
Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA  
Fax: 6177265806  
Email: fishman@gh.cvr.harvard.edu  
http://zebrafish.mgh.harvard.edu  
Primer A: CCTACACGACACTGACCA  
Primer B: TCAAGTCAACGGAAGTCCC  
STS size: 166  
PCR profile:  
Presnak: 94 degrees C for 5.0 minutes  
Denaturation: 94 degrees C for 1.0 minute  
Annealing: 58 degrees C for 1.0 minute  
Polymerization: 72 degrees C for 1.5 minute  
PCR Cycles: 27  
Thermal Cycler: MJ Research PTC-100  
Protocol:  
Template: 10 ng  
Primer: each 375 nm  
dNTPs: each 200 uM  
Tag Polymerase: 0.034 units/uL  
Total Vol: 10 uL

Buffer:  
MgCl2: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3.

FEATURES  
source  
1..279  
/location/Qualifiers  
1..279  
/organism="Danio rerio"  
/strain="AB"  
/db\_xref="taxon:7955"  
/clone="Z15539"  
/clone\_lib="zebrafish AB"  
/sex="F"

/dev\_stage="Adult"  
/lab\_host="DH5alphaF'10"  
/note="Vector: m13mp19 with added BstXI site; V-type:  
Phase: Genomic DNA from a single adult zebrafish of AB  
strain was digested with Alur, Cac81, HaeIII, NlaV, or  
RsaI. Fragments in the range of 250-500 bp were gel  
purified and a BstXI linker was added. The fragments were  
cloned into a modified M13mp19 vector and transformed  
into E. Coli DH5alpha. Microsatellite sequences were  
screened with labeled d(CA)15 and d(GT)15 oligonucleotide  
probes."  
STS  
4. .169  
Primer\_bind 4. .23  
complement(150. .169)  
BASE COUNT 82 a 74 c 38 g 85 t  
ORIGIN

Query Match 7.5%; Score 19; DB 14; Length 279;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 AAGTGACCTCAGAGTA 177  
|||||  
Db 46 AAGTGACCTCAGAGTA 28

RESULT 6  
LOCUS FRTCRAL 24813 bp DNA VRT 05-MAY-1999  
DEFINITION Fugu rubripes T-cell receptor alpha-chain (TCRA) gene, partial  
sequence.  
ACCESSION AF110525  
NID 94741810  
VERSION AF110525.1 GI:4741810  
KEYWORDS  
SEGMENT 1 of 2  
SOURCE Fugu rubripes.  
ORGANISM Fugu rubripes.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorphi;  
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.

REFERENCE 1 (bases 1 to 24813)

AUTHORS Wang, K., Gan, L., Kunisada, T., Lee, I., Yamagishi, H. and Hood, L.  
TITLE Characterization of the Japanese pufferfish T-cell receptor alpha  
locus revealed a unique genomic organization

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 24813)  
AUTHORS Wang, K., Gan, L., Kunisada, T., Lee, I., Yamagishi, H. and Hood, L.  
TITLE Direct Submission  
JOURNAL Submitted (02-DEC-1998) Chiroscience R & D, 1631 220th St. SE,  
Bothell, WA 98021, USA

FEATURES  
Source location/Qualifiers  
1. .24813

J\_segment /organism="Fugu rubripes"  
/db\_xref="taxon:31033"  
/germline  
218. .280  
/gene="TCRA"  
J\_segment /note="15; putative"  
492. .551  
/gene="TCRA"  
J\_segment /note="14; putative"  
1615. .1677  
/gene="TCRA"  
J\_segment /note="13; putative"  
2018. .2075  
/gene="TCRA"  
J\_segment /note="12; putative"  
2413. .2464  
/gene="TCRA"  
J\_segment /note="11; putative"  
2816. .2878

/gene="TCRA"  
/note="10; putative"  
3090. .3149  
/gene="TCRA"  
J\_segment /note="9; putative"  
3593. .3654  
/gene="TCRA"  
J\_segment /note="8; putative"  
3851. .3913  
/gene="TCRA"  
J\_segment /note="7; putative"  
4108. .4168  
/gene="TCRA"  
J\_segment /note="6; putative"  
5000. .5062  
/gene="TCRA"  
J\_segment /note="5; putative"  
5324. .5383  
/gene="TCRA"  
J\_segment /note="4; putative"  
6271. .6333  
/gene="TCRA"  
J\_segment /note="3; putative"  
6589. .6648  
/gene="TCRA"  
J\_segment /note="2; putative"  
7152. .7214  
/gene="TCRA"  
J\_segment /note="1; putative"  
join(8067. .8270,8347. .8391,8472. .8579,8672. .8807)  
/gene="TCRA"  
V\_region complement(join(13412. .13718,13850. .13892))  
/gene="TCRA"  
V\_region complement(join(18244. .18556,18643. .18685))  
/gene="TCRA"  
V\_region complement(join(20980. .21286,21418. .21460))  
/gene="TCRA"  
V\_region complement(join(21767. .22076,22159. .22177))  
/gene="TCRA"

BASE COUNT 7326 a 4969 c 5573 g 6945 t  
ORIGIN

Query Match 7.1%; Score 18; DB 4; Length 24813;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 GCAGCAGCTCAGAGGG 237  
|||||  
Db 16878 GCAGCAGCTCAGAGGG 16861

RESULT 7  
LOCUS FRTCRAL 13801 bp DNA VRT 05-MAY-1999  
DEFINITION Fugu rubripes T-cell receptor alpha-chain (TCRA) gene, partial  
sequence.  
ACCESSION AF110526  
NID 94741811  
VERSION AF110526.1 GI:4741811  
KEYWORDS  
SEGMENT 2 of 2  
SOURCE Fugu rubripes.  
ORGANISM Fugu rubripes.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorphi;  
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.  
REFERENCE 1 (bases 1 to 13801)  
AUTHORS Wang, K., Gan, L., Kunisada, T., Lee, I., Yamagishi, H. and Hood, L.  
TITLE Characterization of the Japanese pufferfish T-cell receptor alpha  
locus revealed a unique genomic organization  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 13801)



```

repeat_region 20990..21290
/Note="AluX repeat: matches 302. .1 of consensus"
repeat_region 22590..22706
/Note="FLAM_C repeat: matches 131. .20 of consensus"
repeat_region 23176..23469
/Note="Aluub repeat: matches 301. .1 of consensus"
repeat_region 23497..23920
/Note="LIME3 repeat: matches 902. .456 of consensus"
repeat_region 24487..24591
/Note="BC200 repeat: matches 1. .107 of consensus"
repeat_region 24603..24730
/Note="Aluio repeat: matches 171. .298 of consensus;
incomplete repeat"
repeat_region 24931..25228
/Note="Aluuo repeat: matches 1. .302 of consensus"
repeat_region 26885..27371
/Note="MLTID repeat: matches 505. .1 of consensus"
repeat_region 27390..27593
/Note="L1 repeat: matches 3702. .3922 of consensus"
repeat_region 27640..27691
/Note="MLT2B repeat: matches 443. .392 of consensus"
repeat_region 27666..27726
/Note="MLT2B repeat: matches 61. .1 of consensus"
repeat_region 27719..27988
/Note="LIME3A repeat: matches 203. .479 of consensus"
repeat_region 27992..28292
/Note="AluSg repeat: matches 1. .301 of consensus"
repeat_region 28337..28503
/Note="LIME8 repeat: matches 515. .677 of consensus"
repeat_region 28884..29185
/Note="AluSx repeat: matches 302. .1 of consensus"
repeat_region 30698..30864
/Note="MIR repeat: matches 86. .262 of consensus"
repeat_region 32237..32421
/Note="MIR repeat: matches 40. .244 of consensus"
repeat_region 32383..32434
/Note="MIR2 repeat: matches 84. .135 of consensus"
repeat_region 33879..34181
/Note="AluSc repeat: matches 299. .1 of consensus"
repeat_region 34182..34304
/Note="MIR repeat: matches 122. .250 of consensus"
repeat_region 34944..35246
/Note="AluY repeat: matches 301. .1 of consensus"
repeat_region 37652..37768
/Note="MIR repeat: matches 47. .166 of consensus"
repeat_region 37888..37981
/Note="MIR2 repeat: matches 58. .145 of consensus"
prim_transcript 39767..40039
/Note="match: 5' ESTS T98231 T98238"
repeat_region 40380..40678
/Note="AluSg repeat: matches 1. .299 of consensus"
repeat_region 41169..41491
/Note="AluY repeat: matches 1. .300 of consensus"
repeat_region 43451..43753
/Note="AluY repeat: matches 2. .301 of consensus"
repeat_region 45053..45148
/Note="6 copies of 16 mer 80 & conserved"
repeat_region 45154..45316
/Note="Aluub repeat: matches 293. .130 of consensus;
incomplete repeat"
repeat_region 45366..45437
/Note="Aluuo repeat: matches 134. .60 of consensus;
incomplete repeat"
repeat_region 45546..45735
/Note="2 copies of 95 mer 84 & conserved"
repeat_region 45638..46069
/Note="27 copies of 16 mer 85 & conserved"
unsure 45652
repeat_region 46069..46203
/Note="Aluio repeat: matches 135. .2 of consensus;
incomplete repeat"
repeat_region 46323..46636
/Note="Aluio repeat: matches 2. .302 of consensus"

```

```

repeat_region 47650..47952
/Note="AluX repeat: matches 1. .302 of consensus"
repeat_region 48817..49026
/Note="MIR repeat: matches 11. .262 of consensus"
repeat_region 50229..50276
/Note="24 copies of 2 mer 83 & conserved"
repeat_region 50277..50546
/Note="AluX repeat: matches 302. .37 of consensus;
incomplete repeat"
repeat_region 51505..51800
/Note="AluSp repeat: matches 1. .298 of consensus"
repeat_region 52529..52826
/Note="AluSg repeat: matches 2. .299 of consensus"
repeat_region 52923..53103
/Note="MIR repeat: matches 3. .179 of consensus"
repeat_region 56253..56586
/Note="Aluio repeat: matches 1. .300 of consensus"
repeat_region 56803..57110
/Note="AluSp repeat: matches 302. .1 of consensus"
repeat_region 57150..57449
/Note="AluSx repeat: matches 302. .2 of consensus"
repeat_region 58140..58321

```

```

Query Match 7.1% Score 18; DB 9; Length 94319;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 146 CTGCCATGTACAAAGTG 163
DB 35443 CTGCCATGTACAAAGTG 35460
|||||

```

```

RESULT 9
LOCUS HSAC002089
DEFINITION Human BAC clone RG308B22 from 7q22-q31, complete sequence.
ACCESSION AC002089
VERSION 92085776
KEYWORDS AC002089.1 GI:2085776
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS Tin-Wollam,A, Graves,T, Sutterer,C and Hawkins,M.
TITLE The sequence of H. sapiens BAC clone RG308B22
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 148416)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-1997)
COMMENT SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:saplens@watson.wustl.edu

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR/> or send an E-mail to [egreen@nhgri.nih.gov](mailto:egreen@nhgri.nih.gov)

**SOURCE INFORMATION:**

This clone is from a release of the human BAC library. The library contains cloned DNA from a human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pRelBAC11

Selection: chloramphenicol

**NEIGHBORING SEQUENCE INFORMATION:**

Actual start of this clone is at base position 1 of H\_RG308B22; actual end is at 148416 of H\_RG308B22. The orientation of this cosmid is unknown.

This clone contains the STS SWSS3356 (NID:g1113751), SWSS2791 (NID:g113586).

```
FEATURES
source          Location/Qualifiers
                1..148416 Homo sapiens
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="7"
                /clone="RG308B22"
                /clone_1lb="CITB-HS-A"
                /map="7q22-q31"
                1..409
                /rpt_family="L1"
                complement(1802..2193)
                /rpt_family="L1"
                5307..5600
                /rpt_family="ALU"
                5991..6019
                /rpt_family="L1"
                complement(9603..9672)
                /rpt_family="L1"
                complement(9923..10112)
                /rpt_family="ALU"
                complement(11034..11056)
                /rpt_family="L1"
                13099..14075
                /rpt_family="L1"
                14076..14366
                /rpt_family="ALU"
                14368..16052
                /rpt_family="L1"
                16083..16665
                /rpt_family="L1"
                16705..17520
                /rpt_family="L1"
                17643..17675
                /rpt_family="L1"
                complement(18933..18957)
                /rpt_family="L1"
                19976..20264
                /rpt_family="ALU"
                21118..21234
                /rpt_family="ALU"
                21261..21433
                /rpt_family="ALU"
                23244..23269
                /rpt_family="L1"
                25137..25166
                /rpt_family="L1"
                complement(27327..27580)
                /rpt_family="MER"
                27705..27876
                /note="similar to human EST H89084 (NID:g1071344)"
```

```
repeat_region  yw24h05.r1
                /rpt_family="L1"
                27877..28462
                /rpt_family="L1"
                28506..28559
                /rpt_family="L1"
                29065..32869
                /rpt_family="L1"
                complement(31746..32160)
                /rpt_family="L1"
                33008..33301
                /rpt_family="L1"
                complement(37975..38010)
                /rpt_family="L1"
                38967..38992
                /rpt_family="L1"
                complement(41143..41378)
                /rpt_family="ALU"
                complement(41399..41457)
                /rpt_family="ALU"
                complement(42386..42517)
                /rpt_family="L1"
                42530..42793
                /rpt_family="ALU"
                complement(42796..43293)
                /rpt_family="L1"
                48888..48930
                /rpt_family="L1"
                complement(50562..50582)
                /rpt_family="L1"
                52062..52102
                /rpt_family="L1"
                54276..54295
                /rpt_family="L1"
                complement(56457..56574)
                /rpt_family="ALU"
                complement(56895..56914)
                /rpt_family="L1"
                complement(58511..58559)
                /rpt_family="L1"
                60810..60836
                /rpt_family="L1"
                complement(62605..62891)
                /rpt_family="ALU"
                complement(63415..63705)
                /rpt_family="ALU"
                64947..64966
                /rpt_family="L1"
                67032..67070
                /rpt_family="L1"
                67536..67643
                /rpt_family="ALU"
                complement(71094..71120)
                /rpt_family="L1"
                complement(71122..71414)
                /rpt_family="ALU"
                72968..73113
                /rpt_family="L1"
                73200..73814
                /rpt_family="L1"
                74806..74847
                /rpt_family="L1"
                complement(75315..75362)
                /rpt_family="L1"
                75547..75824
                /rpt_family="ALU"
                76075..76100
                /rpt_family="L1"
                complement(76956..77010)
                /rpt_family="MER"
                complement(77217..77545)
                /rpt_family="MER"
                77762..78101
                /rpt_family="MER"
                repeat_region
```



repeat\_region complement(79268..79361)  
/rpt\_family="LI"  
repeat\_region complement(80586..80630)  
/rpt\_family="LI"  
repeat\_region complement(80711..82017)  
/rpt\_family="LI"  
repeat\_region 81058..81101  
/rpt\_family="LI"  
repeat\_region complement(82447..82489)  
/rpt\_family="LI"  
repeat\_region complement(82865..83157)  
/rpt\_family="ALU"  
repeat\_region complement(83821..84018)  
/rpt\_family="ALU"  
repeat\_region 84792..84820

Query Match 7.1%; Score 18; DB 10; Length 148416;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 AAAAGATTACCTGCAGCA 225  
|||||  
DB 74072 AAAAGATTACCTGCAGCA 74089

RESULT 10  
AC002460 AC002460 174952 bp DNA PRI 20-AUG-1997  
LOCUS Human BAC clone RG191D16, complete sequence.  
DEFINITION AC002460  
AC002460  
AC002460  
AC002460  
AC002460.1 GI:2337869  
VERSION AC002460.1 GI:2337869  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 174952)  
AUTHORS Zhu,H, Graves,T, Duckels,G and Morris,M.  
TITLE The sequence of H. sapiens BAC clone RG191D16  
JOURNAL Unpublished (1997)  
REFERENCE 2 (bases 1 to 174952)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (20-AUG-1997) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT  
SUBMITTED BY: Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:sapiens@wustl.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

NEIGHBORING SEQUENCE INFORMATION:  
The actual start of this clone is at base position 1 of H\_RG191D16;  
the actual end is at base position 174952 of H\_RG191D16. The orientation of this clone is unknown.

The location of this clone is unknown.

FEATURES

Location/Qualifiers

source 1..174952  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RG191D16"  
/clone\_1b="CITB-978SK-b"  
repeat\_region 1607..1898  
/rpt\_family="ALU"  
repeat\_region 3884..3928  
/rpt\_family="LI"  
repeat\_region 7433..7463  
/rpt\_family="LI"  
repeat\_region 8681..8951  
/rpt\_family="ALU"  
repeat\_region complement(12699..13000)  
/rpt\_family="ALU"  
repeat\_region complement(16803..17005)  
/rpt\_family="ALU"  
repeat\_region 17394..17503  
/rpt\_family="ALU"  
repeat\_region complement(18086..18315)  
/rpt\_family="LI"  
repeat\_region 20793..21031  
/rpt\_family="LI"  
repeat\_region 21726..21783  
/rpt\_family="LI"  
repeat\_region complement(22034..22340)  
/rpt\_family="ALU"  
repeat\_region 23728..23783  
/rpt\_family="LI"  
repeat\_region 24379..24671  
/rpt\_family="ALU"  
repeat\_region 24674..24694  
/rpt\_family="LI"  
repeat\_region 25513..25636  
/rpt\_family="ALU"  
repeat\_region complement(25665..25955)  
/rpt\_family="ALU"  
repeat\_region 26090..26124  
/rpt\_family="LI"  
repeat\_region 28804..28837  
/rpt\_family="LI"  
repeat\_region complement(28992..29282)  
/rpt\_family="ALU"  
repeat\_region 32707..33285  
/rpt\_family="LI"  
repeat\_region 33287..33578  
/rpt\_family="ALU"  
repeat\_region 33582..33624  
/rpt\_family="LI"  
repeat\_region 35052..35156  
/rpt\_family="ALU"  
repeat\_region 35701..35722  
/rpt\_family="LI"  
repeat\_region 36248..36273  
/rpt\_family="LI"  
repeat\_region complement(36446..36590)  
/rpt\_family="ALU"  
repeat\_region 36911..37186  
/rpt\_family="ALU"  
repeat\_region complement(37668..38613)  
/rpt\_family="LI"  
repeat\_region 40477..40555  
/rpt\_family="LI"  
repeat\_region complement(41025..42556)  
/rpt\_family="THR"  
repeat\_region 46364..46383  
/rpt\_family="LI"  
repeat\_region complement(49232..49520)  
/rpt\_family="ALU"  
repeat\_region complement(49631..50045)  
/rpt\_family="THR"  
repeat\_region 52637..52658  
/rpt\_family="LI"

```

repeat_region complement(52928. .53217)
/rpt_family="ALU"
repeat_region complement(53231. .53307)
/rpt_family="LI"
repeat_region 54956. .54975
/rpt_family="LI"
repeat_region complement(55163. .55272)
/rpt_family="ALU"
repeat_region 55276. .55395
/rpt_family="ALU"
repeat_region 55420. .55553
/rpt_family="ALU"
repeat_region complement(57163. .57227)
/rpt_family="LI"
repeat_region 58764. .58831
/rpt_family="ALU"
repeat_region complement(58863. .60449)
/rpt_family="THR"
repeat_region 60600. .60813
/rpt_family="ALU"
repeat_region complement(62252. .62279)
/rpt_family="LI"
repeat_region 63986. .64126
/rpt_family="ALU"
repeat_region complement(68408. .68427)
/rpt_family="THR"
repeat_region 69573. .69866
/rpt_family="ALU"
repeat_region complement(72827. .72857)
/rpt_family="THR"
repeat_region complement(76495. .76514)
/rpt_family="LI"
repeat_region 81693. .81720
/rpt_family="LI"
repeat_region complement(81765. .81966)
/rpt_family="LI"
repeat_region complement(82782. .82804)
/rpt_family="LI"
repeat_region 83279. .83327
/rpt_family="LI"
repeat_region 90566. .95383
/rpt_family="LI"
repeat_region 95428. .95847
/rpt_family="LI"
repeat_region 95858. .95903
/rpt_family="ALU"
repeat_region 97129. .97217
/rpt_family="LI"
repeat_region complement(98666. .98687)
/rpt_family="LI"
repeat_region complement(99082. .99275)
/rpt_family="LI"
repeat_region 99308. .99352
/rpt_family="LI"
repeat_region complement(99664. .99729)
/rpt_family="LI"
repeat_region complement(100135. .100364)
/rpt_family="ALU"
repeat_region complement(104033. .104323)
/rpt_family="LI"
repeat_region 109132. .109178
/rpt_family="LI"
repeat_region complement(114621. .114855)
/rpt_family="ALU"
repeat_region 119374. .119435
/rpt_family="LI"
repeat_region 124463. .124519
/rpt_family="LI"
repeat_region complement(124730. .124983)
/rpt_family="ALU"
repeat_region 128258. .128262
/rpt_family="LI"
repeat_region 129641. .129935

```

```

repeat_region /rpt_family="ALU"
130489. .130529
/rpt_family="LI"
repeat_region complement(131921. .132212)
/rpt_family="ALU"
repeat_region 134621. .134656
/rpt_family="LI"
repeat_region complement(135043. .135213)

Query Match 7 1% Score 18; DB 11; Length 174952;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 TGGGATGTGACATTGA 140
Db 48936 TGGGATGTGACATTGA 48953

RESULT 11
AC002523/C AC002523 228516 bp DNA PRI 05-MAY-1998
LOCUS Homo sapiens Xq28 BACs 360 F12, GSHB-555C13, complete sequence.
DEFINITION AC002523 AC002374 AC002373
ACCESSION 92815518
VERSION AC002523.1 GI:2815518
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 228516)
Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Haywood,M.,
Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L., Shen,H.,
Worley,K., Chen,E., Forcum,J., Arenson,A.D., Gorrell,J.H.,
Brundage,E., Di,W., Chinnault,C., Nelson,D. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 228516)
Chiu,M.W.
Direct Submission
Submitted (10-SEP-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 228516)
Worley,K.C.
Direct Submission
Submitted (28-JAN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 228516)
Worley,K.C.
Direct Submission
Submitted (05-MAY-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.

Beginning of sequence overlaps with AF007262, end of sequence
overlaps with AF011889.
Location/Qualifiers
1. .228516
FEATURES
source

```

```
misc_feature
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
1.149
/note="Positions 104645 to 104793 in AF007262 overlap with
positions 1-149 in this clone"
/note="Region: Overlap with AF007262"

repeat_region
/rpt_family="MER5B"
complement(1270..1577)

repeat_region
/rpt_family="MLT1E"
complement(1653..1708)

repeat_region
/rpt_family="MLT1D"
complement(2282..2306)

repeat_region
/rpt_family="AT_r1ch"
3316..3617

repeat_region
/rpt_family="MLT1B"
3618..3651

repeat_region
/rpt_family="(CA)n"
complement(3658..3944)

repeat_region
/rpt_family="Alusx"
3949..4012

repeat_region
/rpt_family="MLT1B"
complement(4098..4386)

repeat_region
/rpt_family="LINE2"
4723..4784

repeat_region
/rpt_family="MIR"
5630..5727

repeat_region
/rpt_family="(GA)n"
7337..7584

repeat_region
/rpt_family="LINE2"
7678..7709

repeat_region
/rpt_family="(CA)n"
7710..7863

repeat_region
/rpt_family="LINE2"
complement(7981..8151)

repeat_region
/rpt_family="LINE2"
complement(8177..8274)

repeat_region
/rpt_family="LINE2"
complement(10118..10192)

repeat_region
/rpt_family="LINE2"
complement(10400..10443)

repeat_region
/rpt_family="AT_r1ch"
complement(10668..10695)

repeat_region
/rpt_family="AT_r1ch"
11811..11896

repeat_region
/rpt_family="LINE2"
12526..12605

repeat_region
/rpt_family="LINE2"
14892..14932

repeat_region
/rpt_family="MLT1A2"
15361..15661

repeat_region
/rpt_family="Alusp"
16706..16765

repeat_region
/rpt_family="AT_r1ch"
17193..17764

repeat_region
/rpt_family="L1PA7"
17811..18912

repeat_region
/rpt_family="L1PA7"
complement(18968..19018)

repeat_region
/rpt_family="L1"
19034..19085

repeat_region
/rpt_family="L1"
complement(19084..19727)

repeat_region
/rpt_family="L1"
19737..19971

repeat_region
/rpt_family="L1"
19971..20406

repeat_region
/rpt_family="L1ME3A"
20452..20500

repeat_region
/rpt_family="AT_r1ch"
20899..21008

/

repeat_region
/rpt_family="(CAAA)n"
21061..21098

repeat_region
/rpt_family="(TA)n"
complement(21283..21343)

repeat_region
/rpt_family="AT_r1ch"
complement(21527..22106)

repeat_region
/rpt_family="MLT1F"
complement(22286..22326)

repeat_region
/rpt_family="AT_r1ch"
23171..23316

repeat_region
/rpt_family="MER44A"
23311..23626

repeat_region
/rpt_family="MER44C"
complement(23681..23757)

repeat_region
/rpt_family="AT_r1ch"
24613..24924

repeat_region
/rpt_family="Alusx"
complement(24927..25278)

repeat_region
/rpt_family="MSR1A"
25554..25668

repeat_region
/rpt_family="MIR"
complement(27230..27359)

repeat_region
/rpt_family="MER45"
complement(28025..28081)

repeat_region
/rpt_family="MER45"
28227..28312

repeat_region
/rpt_family="MIR"
complement(28521..28607)

repeat_region
/rpt_family="MIR"
complement(29882..29953)

repeat_region
/rpt_family="AT_r1ch"
29960..30154

repeat_region
/rpt_family="MIR"
complement(30336..30383)

repeat_region
/rpt_family="AT_r1ch"
complement(31519..31757)

repeat_region
/rpt_family="MIR"
32198..32905

repeat_region
/rpt_family="L1MA9"
33664..33841

repeat_region
/rpt_family="MIR"
complement(34528..34559)

repeat_region
/rpt_family="AT_r1ch"
complement(34576..34659)

repeat_region
/rpt_family="MIR"
complement(34735..35097)

repeat_region
/rpt_family="MSYD"
complement(35249..35546)

repeat_region
/rpt_family="Aluy"
complement(36693..36767)

repeat_region
/rpt_family="(CA)n"
complement(37299..37786)

repeat_region
/rpt_family="MLT1D"
38010..38096

SRS
/note="147526, HDMSX2596, Chr. X"
/db_xref="dbSTS:17731"

repeat_region
38232..38255

repeat_region
/rpt_family="AT_r1ch"
39138..39218

repeat_region
/rpt_family="MIR"
complement(39901..40077)

repeat_region
/rpt_family="L1MC3"
complement(40079..40208)

repeat_region
/rpt_family="AlusG/x"
complement(40209..40353)

repeat_region
/rpt_family="L1MC3"
complement(40354..40646)

repeat_region
/rpt_family="MER47B"
complement(40647..40798)

repeat_region
/rpt_family="L1MC3"
complement(40793..42299)

repeat_region
/rpt_family="L1MC3"
complement(42297..42391)
```

/rpt\_family="L1MD3"

Query Match 7.1%; Score 18; DB 11; Length 228516;

Best Local Similarity 100.0%; Pred. No. 9.4;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 AGATTACTGCAGCAGCT 228

DB 112027 AGATTACTGCAGCAGCT 112010

## RESULT 12

AC005296

LOCUS

AC005296 192548 bp DNA PRI 26-NOV-1998

DEFINITION

Homo sapiens Xp22 Bins 35-37 BAC GSHB-214D18 (Genome Systems Human

AC005296

AC005296

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: [http://gc.bcm.tmc.edu:8088/quality\\_info/genbank.annotation.html](http://gc.bcm.tmc.edu:8088/quality_info/genbank.annotation.html).

## QUALSTAT-REPORT

FEATURES	Location/Qualifiers
source	1. 192548
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="GSHB-214D18"
	/chromosome="Xp22-35-37"
STS	24704 . 24994
	/standard_name="DXF22S1"
STS	25580 . 25844
	/db_xref="GDB:189989"
	/standard_name="DXS9982"
STS	35126 . 35387
	/db_xref="GDB:9834355"
	/standard_name="DXF22S1"
	/db_xref="GDB:293967"
gene	94713 . 94778 . 104665 . 104907 . 105989 . 108024
	/standard_name="DXS9745"
STS	106347 . 106491
	/db_xref="GDB:293967"
	/standard_name="DXS9745"
STS	107092 . 107240
	/db_xref="GDB:293967"
	/standard_name="DXS9745"
misc_feature	162435 . 163435
	/db_xref="GDB:270551"
	/standard_name="DXS1283E"
BASE COUNT	58221 a 38161 c 37057 g 58109 t 1000 others
ORIGIN	

Query Match 7.1%; Score 18; DB 11; Length 192548;  
Best Local Similarity 100.0%; Pred. No. 9.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 TTACCTGCAGCAGCTTCA 231

DB 22408 TTACCTGCAGCAGCTTCA 22425

## RESULT 13

AC005611/C

LOCUS

DEFINITION

AC005611

AC005611

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

```
REFERENCE 2 (bases 1 to 107190)
AUTHORS  Rickx,D.O.
TITLE      Large Scale Sequence Analysis and Annotation with the Sequence
           Comparison Analysis (SCAN) System
JOURNAL    Unpublished
REFERENCE  3 (bases 1 to 107190)
AUTHORS    Kimmery,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
           Kader,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
           Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE      Direct Submission
JOURNAL    Submitted (04-SEP-1998) Human Genome Center, DOE Joint Genome
           Institute, Lawrence Berkeley National Laboratory, MS 74-157,
           Berkeley, CA 94720, U.S.A.
COMMENT    Sequence submitted by:
           DOE Joint Genome Institute.
           Location/Qualifiers
FEATURES   1. 107190
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="5q"
            /clone="BAC 259n9"
            /chromosome="5"
            /note="LBNI H193"
            repeat_region 2606..2659
            /rpt_family="L1"
            repeat_region 3102..3361
            /rpt_family="Alu"
            repeat_region 3387..3454
            /note="(AAG)17"
            /rpt_type=tandem
            /rpt_unit=AAG
            3455..3514
            /note="(AAGG)15"
            /rpt_type=tandem
            /rpt_unit=AAGG
            5767..6017
            /rpt_family="L1"
            misc_feature 6381..6463
            /note="GRAIL 2 excellent exon, frame 1"
            repeat_region 7397..7566
            /rpt_family="MER3"
            complement(9631..9922)
            /rpt_family="Alu"
            repeat_region 10428..10699
            /rpt_family="Alu"
            repeat_region 13258..13622
            /rpt_family="THE1"
            repeat_region 14070..14340
            /rpt_family="Alu"
            repeat_region 15625..15668
            /note="(GT)22"
            /rpt_type=tandem
            /rpt_unit=GT
            complement(21904..22063)
            /rpt_family="MIR"
            repeat_region 22583..22868
            /rpt_family="Alu"
            repeat_region 23545..23937
            /rpt_family="MLT1"
            repeat_region 25020..25298
            /rpt_family="Alu"
            repeat_region complement(28648..28811)
            /rpt_family="MER20"
            repeat_region 29399..29459
            /rpt_family="MER42"
            complement(29905..30197)
            /rpt_family="LIR10"
            repeat_region complement(30967..31222)
            /rpt_family="Alu"
            repeat_region complement(31367..31651)
            /rpt_family="Alu"
            complement(32445..32727)
            /rpt_family="Alu"

misc_feature 33096..33214
            /note="GRAIL 2 excellent exon, frame 0"
            repeat_region complement(33888..34006)
            /rpt_family="L1"
            repeat_region 36056..36340
            /rpt_family="Alu"
            repeat_region 36323..36344
            /note="(A)22"
            /rpt_type=tandem
            /rpt_unit=A
            repeat_region complement(36641..36716)
            /rpt_family="L1"
            repeat_region complement(36892..37173)
            /rpt_family="Alu"
            repeat_region complement(40185..40391)
            /rpt_family="Alu"
            repeat_region complement(40667..40984)
            /rpt_family="Alu"
            42184..42351
            /note="GRAIL 2 excellent exon, frame 2"
            repeat_region complement(49482..53596)
            /rpt_family="L1"
            repeat_region complement(52004..52222)
            /rpt_family="MER25"
            53613..54359
            /rpt_family="L1"
            repeat_region 59115..59262
            /rpt_family="MER5"
            complement(67368..67655)
            /rpt_family="Alu"
            repeat_region complement(68310..68487)
            /rpt_family="Alu"
            complement(75866..75981)
            /note="GRAIL 2 excellent exon, frame 0"
            78431..79189
            /rpt_family="LIR12"
            complement(83035..83279)
            /rpt_family="Alu"
            repeat_region 84451..84731
            /rpt_family="Alu"
            repeat_region 86146..87094
            /rpt_family="L1"
            repeat_region 87220..87480
            /rpt_family="Alu"
            repeat_region 87797..87935
            /rpt_family="L1"
            repeat_region 87997..88220
            /rpt_family="Alu"
            repeat_region 88261..89245
            /rpt_family="L1"
            complement(90459..90768)
            /rpt_family="Alu"
            repeat_region 90780..91827
            /rpt_family="MER42"
            92399..92877
            /rpt_family="MLT1"
            complement(93197..93343)
            /rpt_family="Alu"
            93366..93424
            /note="GRAIL 2 excellent exon, frame 1"
            94229..94373
            /note="GRAIL 2 excellent exon, frame 2"
            96647..96749
            /note="GRAIL 2 excellent exon, frame 2"
            97013..97413
            /rpt_family="MER21"
            100990..101362
            /rpt_family="Alu"
            complement(101585..101866)
            /rpt_family="Alu"
            102900..103199
            /rpt_family="Alu"

BASE COUNT 32925 a 20084 c 20921 g 33260 t
```



```
/gene="gob-5"  
/codon_start=1  
/protein_id="BAA33743.1"  
/db_xref="pid:d1034712"  
/db_xref="pid:g3721912"  
/db_xref="gi:3721912"  
/translation="MESLKSPEVLLILHLSEVLSESLIOLNNGVEGIVTADHDVP  
EDEALIQHKDMVTQASPLYLEPATGRPFYFKNVAILIPESMKAKPEYTRPKLETFRNA  
DVLVSTISPLGNDPEYTHIGACGKGIHILTPDLGKKLTQYGPQDRFVHEMAH  
FRWGVEFNEYNDDEKFFYLSKGRQAVRCSAITGRNQVRCCOGSCITNGKCVIDRYTG  
LYKDNCFVDPDPHONERKASIMENONINSVEFCTEKNHNOEAPNDONORCNLSYWEV  
IOESDFKOTPTMTQPPAPTEPSILQIGRIYCLVDKSGSWLNDRLRBMQASRLF  
LIQTVESGSPWGMWTFPSAAVVOSEFKOLNSGADBDLITKHLPTVSAGTSTSGSLRT  
APVYIKKKYPTIDGSEIVLLTDGEDNTISSCPDLYKSGAIIHTVALGPRAAKELQOLS  
KMTGGLOQYSSDQVONNGLVDAFALSSGNALAQHSIQLESRGVNLQNNOMNGSVI  
VDSVSGKDTLEFLITWTHPTLFIWDPGVEQNGFILDITTEKVALQVPGTAKVGFWK  
YSIQASQDTLTLVTSRAASATLPIITVTPVYKNTGKFPSPFTVYASIROGASPILR  
ASVTALIESVNGKVTLELDNGAGADATKNDGYRSRPTAFDANGRYSVKIMALGV  
TSDRQRAAPRKNRMYIDGMTEDGEVRMNPPEPESYODKQICFSRTSSGGSFYATN  
VPAAPITPDLFPPCOITIDLKASIQCONLVNLTWTPAGDDYDHGRASNTIIRKSTISIVD  
LRDHFNTSLQVNTGTLIPKASSEELFEFEELGNTFGNGTDIFIAIQAVDNKSLKSEI  
SNIAVSVYFIQAQEPPIPEDSDYPPCPDISINSTIPGIHYLKIIMRTWLGEWOVTLGLH"
```

```
BASE COUNT      860 a      718 c      693 g      666 t  
ORIGIN  
  
Query Match      7.1%; Score 18; DB 12; Length 2937;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY      84 GCTTTCTGCTGCAGAC 101  
        |||  
DB      1010 GCTTTCTGCTGCAGAC 1027
```

Search completed: August 6, 1999, 11:35:28  
Job time: 6214 sec

• • • • •

**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:39:59 ; Search time 650.92 Seconds

(without alignments)  
96.860 Million cell updates/sec

Title: US-09-049-696-6

Perfect score: 252

Sequence: 1 CAAAGATGTGTGTTAGT.....GAGGACGTCACCTGCAGC 252

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	6.7	4529	1	Q46121	Porcine sodium ion
2	16	6.3	11601	1	Q13608	ACV synthetase gen
3	16	6.3	13058	1	Q48231	Vector containing
4	16	6.3	4281	1	Q53471	e1x cDNA. Expressi
5	16	6.3	587	1	Q78965	Human immunoglobul
6	16	6.3	13807	1	Q91438	Polycystic kidney
7	16	6.3	14148	1	T13821	Polycystic kidney
8	16	6.3	13807	1	T08807	Polycystic kidney
9	16	6.3	53577	1	T18551	Human polycystic k
10	16	6.3	1096	1	T61588	Human CD4 gene pro
11	16	6.3	962	1	V01674	Connexin-32 encodi
12	16	6.3	14060	1	T94102	Human PDI cDNA. H
13	16	6.3	53577	1	T94108	Human PDI locus b
14	16	6.3	53526	1	T94101	Human PDI gene. H
15	16	6.3	137507	1	V19941	KSHV long unique c
16	16	6.3	1408	1	V29028	Nucleotide sequenc
17	16	6.3	1401	1	V17721	Human T1 receptor
18	16	6.3	33	1	V17724	T1 receptor-like 1
19	16	6.3	926	1	V32989	Human p24 vesicle
20	16	6.3	26338	1	V62134	HSV-2 strain SB5 C
21	16	6.3	20387	1	V62159	HSV-2 strain SB5 C
22	16	6.3	117213	1	V62176	HSV-2 strain SB5 C
23	16	6.3	2715	1	V64624	P. putida cis/cran
24	16	6.3	32207	1	V73805	KSHV LTR DNA (nucl
25	16	6.3	400	1	V77202	Staphylococcus aur
26	16	6.3	29555	1	V74517	Staphylococcus aur
27	16	6.3	2483	1	X04125	Human factor B enc
28	16	6.3	12838	1	X20509	Polynucleotide seq
29	16	6.3	796	1	X51708	DNA encoding a hum
30	16	6.0	260	1	N60215	Sequence of 3'-end
31	15	6.0	12364	1	Q13547	ACVS gene pcbaB. M
32	15	6.0	12364	1	Q13607	ACV synthetase gen
33	15	6.0	1384	1	N50348	Monkey erythropole
34	15	6.0	2786	1	Q22768	Nucleotides 3392-6
35	15	6.0	1910	1	Q34549	Streptomyces livid
36	15	6.0	245	1	Q77140	Human genome fragm
37	15	6.0	1682	1	Q73501	DNA encoding pseud
38	15	6.0	1255	1	T13981	Human body weight
39	15	6.0	2520	1	T09941	Human earl gene p
40	15	6.0	337	1	T26488	Human gene signatu
41	15	6.0	909	1	T36878	741F8 anti-c-erbB-
42	15	6.0	1743	1	T48232	Neuronal nicotinic
43	15	6.0	1698	1	T48233	Nicotinic acetylch

## ALIGNMENTS

C	44	15	6.0	909	1	T78879	Single chain antib
C	45	15	6.0	643	1	X20436	Human secreted pro
ALIGNMENTS							
RESULT 1							
ID	046121	standard; cDNA; 4529 BP.					
AC	046121;						
DE	21-FEB-1994	(first entry)					
DE	Porcine sodium ion/glucose co-transporter beta-subunit gene.						
KM	Sodium ion-glucose co-transporter; beta-subunit; diabetes;						
KW	glucose absorption; insulin demand; reduction; ss.						
OS	Sus scrofa domestica.						
FT	Key	Location/Qualifiers					
FT	cds	2..1570					
FT		/tag- a					
FT		/note- "putative initiator methionine at position					
FT		125-127 within the open reading frame"					
FT		1861..1867					
FT	polya_signal	/tag- b					
FT		/note- "putative"					
FT		2922..2927					
FT		/tag- c					
FT		/note- "putative"					
FT	polya_signal						
FT							
PN	DE4218669-C.						
PD	02-SEP-1993.						
PF	05-JUN-1992; 218669.						
PR	05-JUN-1992; DE-218669.						
PA	(PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.						
PI	Koepsell H;						
DR	WPI; 93-273987/35.						
DR	P-PSDB; R41445.						
PT	New beta sub-unit of sodium-glucose co-transporter - and DNA						
PT	encoding it, useful for regulating glucose absorption and						
PS	excretion, esp. in diabetes						
PS	Claim 1; Fig 6; 18pp; German.						
CC	A pig renal cortex cDNA bank was screened with antibody R446						
CC	directed against porcine sodium ion-glucose cotransporter. Positive						
CC	clone p20 containing a 4.5kb insert was isolated. The insert						
CC	encodes a single polypeptide of mol.wt. 70000. Using p20 cDNA as						
CC	template, cRNA was transcribed and found to significantly increase						
CC	glucose uptake when injected into Xenopus oocytes of low endogenous						
CC	transport activity. The porcine gene sequence was also used to						
CC	design PCR primers for isolating the human sodium-glucose						
CC	cotransporter beta-subunit. See Q46122 and Q46123.						
SQ	Sequence 4529 BP; 1134 A; 1047 C; 1058 G; 1290 T;						
Query Match							
Best local similarity 100.0%; Pred. No. 8.4;							
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
QY	86	TTTTCTGCTGCAGACA	102				
DB	1429	TTTTCTGCTGCAGACA	1445				
RESULT 2							
ID	Q13608	standard; DNA; 11601 BP.					
AC	Q13608;						
DE	22-NOV-1991	(first entry)					
DE	ACV synthetase gene, pcbaB.						
KM	Beta lactam antibiotics; penicillin; ss.						
OS	Acetomonium chrysogenum.						
FT	Key	Location/Qualifiers					
FT	cds	388..11526					
FT		/tag- a					
FT		/product- ACV synthase					
PN	EP-445868-A.						

PD 11-SEP-1991. 200423.  
 PF 27-FEB-1991; EP-200423.  
 PR 28-FEB-1990; EP-200475.  
 PR 28-FEB-1990; EP-200488.  
 PR 02-JUL-1990; EP-201768.  
 PR 03-OCT-1990; EP-202628.  
 PR 27-FEB-1991; EP-200423.  
 PA (KONN) GIST-BROCADES NV.  
 PI Veenstra AE, Martin JF, Garcia BD, Gutierrez S, Barredo JL,  
 PI Montenegro PE, Von Doehren H, Palissa H, Van Llamp H;  
 DR WPI: 91-268735/37.  
 DR P-PSDB: R13896.  
 PT DNA encoding amino:adipyl-L-cysteinyl-valine synthetase - used for  
 PT prodn. of the enzyme or enhanced prodn. of new or known  
 PT beta-lactam antibiotic cpds.  
 PS Claim 1; Page 20; 54pp; English.  
 CC The DNA sequence was obtd. from five subclones isolated from a gene  
 CC library of A. chrysoeum C10 (ATCC 48). The gene can be used to  
 CC express the synthetase enzyme which can be used for the prodn. of  
 CC new beta-lactam anti-biotics.  
 CC See also Q13607-Q13612.  
 SQ Sequence 11601 BP; 2537 A; 3566 C; 3158 G; 2340 T;

Query Match 6.3%; Score 16; DB 1; Length 11601;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 CCTGCAGCAGCTTCAG 232  
 DB 9682 CCTGCAGCAGCTTCAG 9667

RESULT 3  
 ID Q48231/c 048231 standard; DNA; 13058 BP.  
 AC Q48231;  
 DT 21-FEB-1994 (first entry)  
 DE Vector containing ACVS gene.  
 KM Delta-(L-alpha-aminoacidipyl)-L-cystinyl-D-valine synthase; ACVS;  
 KM beta-lactam; antibiotic; transformed; cephalosporin; vector; ss.  
 OS Acromonium chrysoeum.  
 FH Key Location/Qualifiers  
 FT cds 1458..12377  
 FT /tag- a  
 FT /product= ACVS  
 PN J05192162-A.  
 PD 03-AUG-1993.  
 PR 25-JUL-1991; 186222.  
 PR 31-JUL-1990; JP-205677.  
 PA (TAKE) TAKEDA CHEM IND LTD.  
 DR WPI: 93-277475/35.  
 DR P-PSDB: R40227.  
 PT DNA coding delta-(L-alpha-amino-adipyl) L-cystinyl D-valine  
 PT synthase - for improved productivity of cephalosporin antibiotics  
 PS Claim 1; Page 9-13; 69pp; Japanese.  
 CC The sequence (Q48231) is of a vector which includes the  
 CC delta-(L-alpha-aminoacidipyl)-L-cystinyl-D-valine synthase gene.  
 CC This sequence was transformed into a host cell to express the ACVS  
 CC product. The protein produced (R40227) was then used to manufacture  
 CC a beta-lactam antibiotic.  
 SQ Sequence 13058 BP; 2860 A; 3915 C; 3560 G; 2723 T;

Query Match 6.3%; Score 16; DB 1; Length 13058;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 CCTGCAGCAGCTTCAG 232  
 DB 10533 CCTGCAGCAGCTTCAG 10518

RESULT 4  
 ID Q53471/c 053471 standard; cDNA; 4281 BP.  
 AC Q53471;  
 DT 16-JUN-1994 (first entry)  
 DE elk cDNA.  
 KM Lambda gt11; expression vector; lambda-BI-Elk; protein tyrosine kinase;  
 KM Elk; BI; Eph; subfamily; receptor-like tyrosine kinase; eph; eck;  
 KM phosphorylation; phosphorylated kinase insert domain; growth factor;  
 KM receptor kinase; platelet-derived growth factor receptor; ss.  
 OS Rattus rattus.  
 FH Key Location/Qualifiers  
 FT cds 367..3321  
 FT /tag- a  
 FT /product= Elk

PN CA2083521-A.  
 PD 01-OCT-1993.  
 PR 23-NOV-1992; 083521.  
 PR 31-MAR-1992; US-861390.  
 PA (MOUN) MOUNT SINAI HOSPITAL CORP.  
 PI Letwin K, Pawson A, Reedijk M;  
 DR WPI: 93-406300/51.  
 DR P-PSDB: R44513.  
 PT Expression of phosphorylated exogenous protein - in host cells  
 PT transformed with two vectors, one for the protein, the other for  
 PT catalytic domain of protein kinase  
 PS Disclosure; Fig 3; 55pp; English.  
 CC This sequence represents the elk cDNA which encodes the protein  
 CC tyrosine kinase, Elk. The Elk gene, BI, encode a protein which is  
 CC a member of the Eph subfamily of protein tyrosine kinases. The Elk  
 CC product is very similar to two other receptor-like tyrosine kinases,  
 CC eph and eck. Lambda-BI-Elk may be used in the production of  
 CC phosphorylated exogenous protein along with a further vector encoding  
 CC the desired exogenous protein. These plasmid may be used to produce  
 CC phosphorylated proteins in host cells which have no intrinsic capacity  
 CC for phosphorylation, eg. bacteria. The system may be used for the  
 CC expression of the phosphorylated kinase insert domain of a growth  
 CC factor receptor kinase eg. platelet-derived growth factor receptor.  
 SQ Sequence 4281 BP; 1075 A; 1170 C; 1121 G; 915 T;

Query Match 6.3%; Score 16; DB 1; Length 4281;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 TTTCCTGCTGCAGACA 102  
 DB 2064 TTTCCTGCTGCAGACA 2049

RESULT 5  
 ID Q78965 standard; DNA; 587 BP.  
 AC Q78965;  
 DT 03-AUG-1995 (first entry)  
 DE Human immunoglobulin VH gene #27.  
 KM Primer; PCR; amplification; human; immunoglobulin; variable; heavy chain;  
 KM cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 65..487  
 FT /tag- a  
 FT /product= human immunoglobulin variable heavy chain  
 FT Intron 111..194  
 FT /tag- b  
 FT misc\_signal 317..319  
 FT /tag- c  
 FT /note= "non-functional terminator codon"  
 FT misc\_signal 488..490  
 FT /tag- d  
 FT /note= "miscellaneous signal, does not conform to  
 FT terminator or splice site sequence"  
 PN M09426895-A.

PD 24-NOV-1994.  
PF 10-MAY-1993; J00603.  
PR 10-MAY-1993; WO-J00603.  
PA (NISH ) JAPAN TOBACCO INC.  
PI Honjo T, Matsuda F.  
DR WPI: 95-006791/01.  
DR P-PSDB: R66319.  
PT DNA fragment comprising human immunoglobulin Vh genes - for the  
PT production of human immunoglobulin in mammalian hosts  
PS Claim 36; Page 66-67; 130pp; Japanese.  
CC A series of genes (Q78339-79002) encoding human immunoglobulin variable  
heavy chains. The genes were isolated and cloned from a series of cosmid  
constructs: Y202; Y103; Y21; Y6; Y24; 3-31; M84; M18 and M131, by PCR  
amplification using primers Q78917-38. The genes are subdivided into 5  
families of Vh genes. The fragments cover a region of 800 kb. The DNA  
fragments were isolated from high molecular weight DNA from human  
placenta. The DNA was partially digested with RsaI restriction enzyme.  
The fragments were separated by gel electrophoresis and 35-45 kb fractions  
were collected. The fragments were ligated with ClaI-digested cosmid  
vector pUB81. The ligation products were in vitro packed and infected  
into E.coli 490a. The fragments were then subcloned by colony  
hybridisation. The Vh genes and the DNA fragments encoding them are  
useful in producing human immunoglobulin in mammalian hosts.  
SQ Sequence 587 BP; 155 A; 163 C; 159 G; 110 T;

Query Match 6.3%; Score 16; DB 1; Length 587;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 GGCAGTGCACAGGACA 199  
DB 557 GGCAGTGCACAGGACA 572

RESULT 6  
Q91438/c  
ID Q91438 standard; cDNA; 13807 BP.  
AC Q91438;  
DT 14-APR-1996 (first entry)  
DE Polycystic kidney disease 1 cDNA (near complete sequence).  
KW Autosomal dominant polycystic kidney disease; ADPKD;  
KW polycystic kidney disease 1 gene; PKD1; diagnostic; gene therapy; ds.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 2..13021 /\*tag- a  
FT CDS WO9518225-A1.  
PN 06-JUN-1995.  
PD 23-DEC-1994; G02822.  
PR 24-DEC-1993; GB-026470.  
PR 14-JUN-1994; GB-011900.  
PA (MED1-) MEDICAL RES COUNCIL.  
PA (UYLE-) RIJDSUNIV LEIDEN.  
PA (UYRO-) UNIV ROTTERDAM ERASMUS.  
PA (UYWA-) UNIV WALES COLLEGE OF MEDICINE.  
PI Breunig MH, Halley DJJ, Harris PC, Hesselting ALW;  
PI Hughes J, Janssen LAJ, Nellist MD, Peral B, Peters DM;  
PI Roelofsma JH, Sampson J, Ward CJ;  
DR WPI: 95-246390/32.  
DR P-PSDB: R75916.  
PT Isolated polycystic kidney disease 1 gene and its mutants - useful  
PT for treatment and diagnosis of autosomal dominant polycystic kidney  
PT disease  
PS Claim 2; Fig 10; 119pp; English.  
CC A polycystic kidney disease 1 (PKD1) cDNA contig (Q91438) relates  
CC to the entire PKD1 gene sequence apart from the extreme 5' end.  
CC The sequence is compiled from a cDNA contig (see Q91339) obt'd. from  
CC chromosome 16, and various other isolated cDNA and genomic DNAs.  
CC A chromosome translocation associated with autosomal dominant  
CC polycystic kidney disease (ADPKD) disrupts the PKD1 candidate region  
CC at 16p13.3. The PKD1 gene codes for a novel protein (R75916).  
CC Screening for normal or mutated PKD1 can be used for diagnostic

CC purposes, and disorders such as ADPKD may be treated or prevented by  
CC PKD1 gene therapy.  
SQ Sequence 13807 BP; 2144 A; 4666 C; 4502 G; 2495 T;

Query Match 6.3%; Score 16; DB 1; Length 13807;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 GTGCAGTGCACAGGACA 197  
DB 780 GTGCAGTGCACAGGACA 765

RESULT 7  
T13821/c  
ID T13821 standard; cDNA; 14148 BP.  
AC T13821;  
DT 02-FEB-1997 (first entry)  
DE Polycystic kidney disease 1 (PKD1) gene.  
KW Adult polycystic kidney disease; APKD; PKD1 gene; diagnosis;  
KW gene therapy; ds.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 212..13120  
FT /\*tag- a  
FT 13081..13082  
FT conflict /\*tag- b  
FT /\*note- "bases 13081-13082 (CC) are not found  
FT in PKD1 partial transcripts, and alter  
FT the reading frame"

WO9534648-A2.  
PN 21-DEC-1995.  
PD 13-JUN-1995; G01386.  
PR 14-JUN-1994; GB-011900.  
PR 23-DEC-1994; WO-G02822.  
PR 13-APR-1995; GB-007766.  
PR 14-APR-1995; US-422582.  
PA (MED1-) MEDICAL RES COUNCIL.  
PA (UYLE-) RIJDSUNIV LEIDEN.  
PA (UYRO-) UNIV ROTTERDAM ERASMUS.  
PA (UYWA-) UNIV WALES COLLEGE OF MEDICINE.  
PI Breunig MH, Halley DJJ, Harris PC, Hesselting ALW;  
PI Hughes J, Janssen LAJ, Nellist MD, Peral B, Peters DM;  
PI Roelofsma JH, Sampson J, Ward CJ;  
DR WPI: 96-049678/05.  
DR P-PSDB: W00870.  
PT Isolated polycystic kidney disease 1 gene and its deletion mutants  
PT - useful in diagnosis and treatment of PKD1-associated disease and  
PT in gene therapy.  
PS Claim 2; Fig 15; 181pp; English.  
CC The human PKD1 gene (T13821), located on chromosome 16, encodes  
CC a polypeptide (W00870) that has a role in the prevention or  
CC suppression of adult polycystic kidney disease (APKD). The  
CC chromosome 16 translocation breakpoint was located in 2 members  
CC of an APKD family and the gene disrupted by the rearrangement was  
CC defined. The discovery of additional mutations of the gene in  
CC other PKD1 patients confirmed the identity of the gene. Full  
CC characterisation of the transcript was difficult because all but  
CC 3.5 kb at the 3' end is encoded by a region that is reiterated  
CC several times elsewhere in the chromosome. The PKD1 gene can  
CC be used in gene therapy, as a probe to detect PKD1-associated  
CC diseases, or to produce the PKD1 polypeptide.  
SQ Sequence 14148 BP; 2166 A; 4814 C; 4641 G; 2527 T;

Query Match 6.3%; Score 16; DB 1; Length 14148;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 GTGCAGTGCACAGGACA 197  
DB 1119 GTGCAGTGCACAGGACA 1104

RESULT 8  
ID T08807/c  
T08807 standard: cDNA: 13807 BP.  
AC T08807;  
DE 02-FEB-1997 (first entry)  
DE Polycystic kidney disease 1 gene (cDNA fragment).  
KW Adult polycystic kidney disease; APKD; PKD1 gene; diagnosis;  
therapy; ds.  
OS Homo sapiens.  
FH key Location/Qualifiers  
FT cds 2..13021  
FT /tag= a  
FT MO9534649-A2.  
FT 21-DEC-1995.  
PF 13-JUN-1995; G01386.  
PR 14-JUN-1994; GB-011900.  
PR 23-DEC-1994; WO-602822.  
PR 13-APR-1995; GB-007766.  
PR 14-APR-1995; US-422582.  
PA (MEDIC-) MEDICAL RES COUNCIL.  
PA (UYLE-) RIKSUNIV LEIDEN.  
PA (UYRO-) UNIV ROTTERDAM ERASMUS.  
PA (UYMA-) UNIV WALES COLLEGE OF MEDICINE.  
PI Breuning MH, Hailey DJJ, Harris PC, Hesselting ALM;  
PI Hughes J, Janssen LAJ, Nellist MD, Peral B, Peters DJM;  
PI Roelofsma JH, Sampson J, Ward CJ;  
DR WPI: 96-049678/05.  
DR P-PSDB: R87539.  
PT Isolated polycystic kidney disease 1 gene and its deletion mutants  
PT - useful in diagnosis and treatment of PKD1-associated disease and  
PT in gene therapy  
PS Claim 4: Fig 10: 181pp; English.  
CC A partial PKD1 cDNA sequence (T08807) incorporates the entire PKD1  
CC gene sequence (see also T13821) apart from its extreme 5' end. The  
CC PKD1 gene is associated with adult polycystic kidney disease (APKD);  
CC deleted forms of the gene have been identified in PKD1 patients.  
CC The partial cDNA fragment can be de utilised in the design of  
CC probes to screen for PKD1 associated disorders, to detect mutant  
CC PKD1 genes, and to detect disease carriers. It may also be  
CC incorporated into a vector and used to produce PKD1 protein (R98738)  
CC in transformed host cells.  
SQ Sequence 13807 BP; 2143 A; 4663 C; 4506 G; 2495 T;

Query Match 6.3%; Score 16; DB 1; Length 13807;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 GTGGCAGTGACAGGGA 197  
DB 780 GTGGCAGTGACAGGGA 765

RESULT 9  
ID T18551/c  
T18551 standard: DNA: 53577 BP.  
AC T18551;  
DE 06-MAY-1997 (first entry)  
DE Human polycystic kidney disease normal PKD1 gene.  
KW Adult onset polycystic kidney disease; APKD; autosomal dominant;  
KW mutant; transversion; transition; deletion; insertion; ds.  
OS Homo sapiens.  
FH key Location/Qualifiers  
FT misc\_feature 4379..5272  
FT /tag= a  
FT /note= "specifically claimed region of intronless  
FT cDNA identified by exon trapping"  
FT replace(50652..50653, cg)  
FT /tag= b  
FT /note= "changes Val codon to Leu codon"  
FT old\_sequence replace(50796..50797, cg)

FT /tag= c  
FT /note= "replaces Val codon by Leu codon"  
FT insert(51827..51828, cc)  
FT /tag= d  
FT /note= "insertion, results in frameshift"

FT old\_sequence  
FT /tag= d  
FT /note= "insertion, results in frameshift"

PN MO9612033-A1.  
PD 25-APR-1996.  
PF 11-OCT-1995; U13357.  
PR 12-OCT-1994; US-323443.  
PR 31-JAN-1995; US-381520.  
PA (IGTC-) IG LAB INC.  
PA (UYUD-) UNIV JOHNS HOPKINS.  
PI Burn TC, Connors TD, Dackowski W, Germino G, Klingner KW;  
PI Landes GM, Qian F;  
DR WPI: 96-222017/22.  
PT Isolated human polycystic kidney disease and its mutants  
PT useful for treatment of polycystic kidney disease and screening for  
PT carriers  
PS Claim 1: Fig 1; 65pp; English.  
CC The present sequence is that of the normal human PKD1 gene from  
CC chromosome 16. Mutations in this gene (e.g. transitions,  
CC transversions, deletions and/or insertions) are associated with  
CC adult-onset polycystic kidney disease (APKD). The PKD1 locus is  
CC GC-rich (62.4%). Comparison of this sequence with a previously  
CC reported partial cDNA sequence revealed differences at three  
CC locations (see features table). The most significant difference is  
CC the presence of two additional cytosine residues on the plus-strand  
CC at position 4566 of the previously reported sequence. The insertion  
CC results in a frame-shift in the predicted protein coding sequence,  
CC leading to replacement of 92 C-terminal amino acids with a novel  
CC 12 amino acid C-terminus. The PKD1 gene contains 23 Alu repeats.  
CC There is a region consisting of 17 tandem copies of a perfect 27 bp  
CC repeat and two large CT-rich regions.  
SQ Sequence 53577 BP; 8495 A; 17681 C; 15785 G; 11616 T;

Query Match 6.3%; Score 16; DB 1; Length 53577;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 GTGGCAGTGACAGGGA 197  
DB 21197 GTGGCAGTGACAGGGA 21182

RESULT 10  
ID T61588  
T61588 standard: DNA: 1096 BP.  
AC T61588;  
DE 05-NOV-1997 (first entry)  
DE Human CD4 gene promoter.  
KW Murine CD4 enhancer; haematopoietic cell; mature T cell; lymphocyte;  
KW gene therapy; graft rejection; guest versus host disease;  
KW viral infection; human immunodeficiency virus; HIV;  
KW autoimmune disease; ds.  
OS Homo sapiens.  
PN MO9704118-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01122.  
PF 17-JUL-1995; FR-008616.  
PA (UYPA-) UNIV CURIE PARIS VI P & M.  
PI Boyer O, Klatzmann D, Salmon P;  
DR WPI: 97-132655/12.  
PT System for expressing heterologous gene or protein exclusively in  
PT mature T cells - includes enhancer from a CD4 gene, also transformed  
PT haematopoietic cells, esp. for selective destruction of activated T  
PT cells  
PS Claim 2: Fig 6; 51pp; French.  
CC A heterologous gene or protein is expressed only in mature T cells  
CC by transducing haematopoietic cells (especially stem cells) with a  
CC recombinant vector which includes all sequences needed for expression  
CC and at least one enhancer from a CD4 gene from the same or different  
CC species. The enhancer confers specific expression in mature (but not

CC immature) T cells. The system is used in gene therapy for selective  
CC destruction of activated T cells, particularly to prevent graft  
CC rejection or guest versus host disease, or to treat or prevent viral  
CC (e.g. HIV) infection, autoimmune disease or primary or secondary  
CC immune deficiencies. The vector preferably contains the human CD4  
CC gene promoter having the present sequence or nucleotides -498 to +16  
CC or -165 to +16 from it. 243 A; 271 C; 300 G; 282 T;  
SQ Sequence 1096 BP;

Query Match 6.3%; Score 16; DB 1; Length 1096;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 178 AACAGTGCAGTGACA 193  
|||||  
DB 883 AACAGTGCAGTGACA 898

RESULT 11  
ID V01674 standard; DNA; 962 BP.  
AC V01674;  
DT 27-MAR-1998 (first entry)  
DE Connexin-32 encoding DNA.  
KW Connexin-32; human; Charcot-Marie-Tooth disease; X-linked;  
KM detection; mutation; ds.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 63..914  
FT FT /\*tag= a  
FT FT /product= "Connexin-32"

PN US5691144-A.  
PD 25-NOV-1997.  
PF 05-JUN-1996; 658469.  
PR 05-JUN-1996; US-658469.  
PA (ATHE-) ATHENA DIAGNOSTICS INC.  
PI Ananth U, Boss MA, Ottaway K;  
DR WPI: 98-017628/02.  
DR P-PSDB; W37469.  
PT Diagnosis of X-linked Charcot-Marie-Tooth disease - by detecting  
PT mutation in connexin-32 gene  
PS Claim 13; Column 11-14; 13pp; English.  
CC A method has been developed for diagnosing X-linked Charcot-Marie-Tooth  
CC disease. The method comprises detecting a mutation in the connexin-32  
CC gene, the mutation being in at least 1 codon selected from 28, 44, 86,  
CC 100, 103, 124, 154, 157, 160, 161, 179, 181, 183, 187, 198, 204, 205,  
CC 219, 230 and 235, where the presence of the mutation in the connexin-32  
CC gene is indicative of X-linked Charcot-Marie-Tooth disease. The present  
CC sequence encodes connexin-32. A method is also described as above except  
CC that the mutation is selected from a change in the nucleic acid sequence  
CC of codon 12 from GGC to GGT (GGC12GGT), ATC28AAC, TGG44ATG, CGG75TGG,  
CC ACC86TCC, CAC100TAC, AAG103GAG, AAG124AG, CGG142CAG, TAA154TAA,  
CC TAC157TGC, TAT160CAT, GCC161GCT, CCG172CTC, TGC199CTG, GTG18ATG,  
CC CGC183TGC, deletion of codon 185, AAA187GAA, TCT198TTT, CTC204GTC,  
CC AAT205AGT, CGC219TGC, CGC307GC and TTC235TGC. The X-linked disease can  
CC be distinguished from other forms of Charcot-Marie-Tooth disease.  
SQ Sequence 962 BP; 188 A; 296 C; 267 G; 211 T;

Query Match 6.3%; Score 16; DB 1; Length 962;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 GTGCAGTGACAGGGA 197  
|||||  
DB 30 GTGCAGTGACAGGGA 45

RESULT 12  
T94102/c  
ID T94102 standard; cDNA; 14060 BP.  
AC T94102;

DT 01-JUN-1998 (first entry)  
DE Human PKD1 cDNA.  
KW Human; polycystic kidney disease 1; PKD1; treatment;  
KM autosomal dominant polycystic kidney disease; APKD; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 135..13043  
FT FT /\*tag= a  
FT FT /product= PKD1\_polypeptide

PN MO9744457-A1.  
PD 27-NOV-1997.  
PF 22-MAY-1997; U08799.  
PR 03-JUN-1996; US-658136.  
PR 24-MAY-1996; US-655360.  
PA (GENZ ) GENZYME CORP.  
PI Burn T, Connors T, Dackowski W, Germino G, Klinger K,  
PI Qian F;  
DR WPI: 98-018511/02.  
DR P-PSDB; W33396.  
PT Human polycystic kidney disease gene, PKD1 - useful to treat and  
PT diagnose human autosomal or adult onset polycystic kidney disease  
PS Claim 4; Pages 119-138; 257pp; English.  
CC The present sequence is the human polycystic kidney disease 1  
CC (PKD1) cDNA. The PKD1 cDNA or polypeptide may be used to treat  
CC autosomal dominant polycystic kidney disease (APKD), and identify  
CC carriers of mutant PKD1 genes, i.e. subjects susceptible to APKD.  
CC Antibodies (Ab) that distinguish between normal and mutant PKD1  
CC sequences can also be used in diagnostic tests. Anti-PKD1 Ab may  
CC also be used to perform subcellular and histochemical localization  
CC studies, and to block the function of PKD1. Ab are also useful in  
CC rational drug design studies to identify and test inhibitors of  
CC PKD1. Sense and antisense sequences derived from the PKD1 gene may  
CC be used for detection and therapy.  
SQ Sequence 14060 BP; 2140 A; 4789 C; 4614 G; 2517 T;

Query Match 6.3%; Score 16; DB 1; Length 14060;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 GTGCAGTGACAGGGA 197  
|||||  
DB 1042 GTGCAGTGACAGGGA 1027

RESULT 13  
T94108/c  
ID T94108 standard; DNA; 53577 BP.  
AC T94108;  
DT 01-JUN-1998 (first entry)  
DE Human PKD1 locus between chromosomal markers ATP1 (ATP6C) and D16S84.  
KW Human; polycystic kidney disease 1; PKD1; treatment;  
KM autosomal dominant polycystic kidney disease; APKD; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 135..13043  
FT FT /\*tag= a  
FT FT /product= PKD1\_polypeptide

CC perform subcellular and histochemical localisation studies, and to  
CC block the function of PKD1. Ab are also useful in rational drug  
CC design studies to identify and test inhibitors of PKD1. Sense and  
CC antisense sequences derived from the PKD1 gene may used for  
CC detection and therapy.  
SQ Sequence 53577 BP; 8495 A; 17684 C; 15762 G; 11616 T;

Query Match 6.3%; Score 16; DB 1; Length 53577;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 GTGGCAGTGACAGGGA 197  
|||||  
DB 21197 GTGGCAGTGACAGGGA 21181

## RESULT 14

T94101/C  
ID T94101 standard; DNA; 53526 BP.  
AC T94101;  
DT 01-JUN-1998 (first entry)  
DE Human PKD1 gene.  
KW Human; polycystic kidney disease 1; PKD1; treatment;  
KW autosomal dominant polycystic kidney disease; APKD; ss.  
OS Homo sapiens.  
PN WO9744457-A1.  
PD 27-NOV-1997.  
PE 22-MAY-1997; U08799.  
PR 03-JUN-1996; US-658136.  
PR 24-MAY-1996; US-655360.  
PA (GENZ) GENZYME CORP.  
PI Burn T, Connors T, Dackowski W, Germino G, Klingner K,  
PI Qian F;  
DR WPI; 98-018511/02.  
PT Human polycystic kidney disease gene, PKD1 - useful to treat and  
PT diagnose human autosomal or adult onset polycystic kidney disease  
PS Claim 2; Pages 90-118; 257pp; English.  
CC The present sequence is the human polycystic kidney disease 1  
CC (PKD1) gene. The PKD1 gene or polypeptide may be used to treat  
CC autosomal dominant polycystic kidney disease (APKD), and identify  
CC carriers of mutant PKD1 genes, i.e. subjects susceptible to APKD.  
CC Antibodies (Ab) that distinguish between normal and mutant PKD1  
CC sequences can also be used in diagnostic tests. Anti-PKD1 Ab may  
CC also be used to perform subcellular and histochemical localisation  
CC studies, and to block the function of PKD1. Ab are also useful in  
CC rational drug design studies to identify and test inhibitors of  
CC PKD1. Sense and antisense sequences derived from the PKD1 gene may  
CC used for detection and therapy.  
SQ Sequence 53526 BP; 8486 A; 17665 C; 15768 G; 11607 T;

Query Match 6.3%; Score 16; DB 1; Length 53526;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 GTGGCAGTGACAGGGA 197  
|||||  
DB 21197 GTGGCAGTGACAGGGA 21182

## RESULT 15

V19941/C  
ID V19941 standard; DNA; 137507 BP.  
AC V19941;  
DT 03-AUG-1998 (first entry)  
DE KSHV long unique coding region and terminal repeat.  
KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;  
KW interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;  
KW complement-binding protein; glycoprotein; capsid protein IV; infection;  
KW immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;  
KW lymphoproliferative disease; leukemia; splenomegaly; mycosis fungoides;  
KW HIV immune status; anti-inflammatory agent; therapy; ds.

OS Kaposi's sarcoma-associated herpes virus.  
FH Key Location/Qualifiers  
FT CDS 1142..2794

FT /tag- a  
FT /product- complement-binding protein  
FT 8699..11236

FT /tag- b  
FT /product- glycoprotein B  
FT complement (17261..17875)

FT /tag- c  
FT /product- interleukin 6  
FT complement (21548..21832)

FT /tag- d  
FT /product- macrophage inflammatory protein II  
FT complement (27137..27424)

FT /tag- e  
FT /product- interferon regulatory factor 1  
FT 28661..29741

FT /tag- f  
FT /product- protein T1.1  
FT complement (58976..60175)

FT /tag- g  
FT /product- glycoprotein M  
FT complement (69412..69915)

FT /tag- h  
FT /product- glycoprotein L  
FT complement (88410..88910)

FT /tag- i  
FT /product- interferon regulatory factor 2  
FT 89600..90541

FT /tag- j  
FT /product- interferon regulatory factor 3  
FT 90173..90643

FT /tag- k  
FT /product- glycoprotein X  
FT complement (93636..94127)

FT /tag- l  
FT /product- interferon regulatory factor 4  
FT complement (111931..112443)

FT /tag- m  
FT /product- capsid protein IV  
FT complement (123808..127296)

FT /tag- n  
FT /product- immediate early protein  
FT WO9804576-A1.

PD 05-FEB-1998.  
PN U13346.  
PR 22-JUL-1997; US-757669.

PR 25-JUL-1996; US-686243.  
PR 25-JUL-1996; US-686349.

PR 25-JUL-1996; US-686350.  
PR 25-JUL-1996; US-687253.

PR 25-JUL-1996; US-688814.  
PR 05-SEP-1996; US-708678.

PR 10-OCT-1996; US-728323.  
PR 13-NOV-1996; US-747887.

PR 13-NOV-1996; US-748640.  
PA (UYCO ) UNIV COLUMBIA NEW YORK.

PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JF;  
DR WPI; 98-130615/12.

PT New nucleic acid encoding Kaposi's sarcoma associated herpes virus  
PT proteins - useful for, e.g. detecting levels of HHV8 in, and  
PT preparation of vaccines for treatment of, HIV patients

PS Example 2; Page 135-203; 230pp; English.  
CC This sequence represents the long unique region and terminal repeat of  
CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known  
CC as human herpes virus 8 (HHV8). This sequence contains the DNAs of the  
CC invention which encode KSHV polypeptides selected from: (a) viral  
CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);  
CC (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;  
CC (d) capsid protein IV encoded by ORF5; and (e) immediate early protein  
CC encoded by ORF7. Labelled probes for the nucleic acid, proteins encoded  
CC by it, and antibodies (Ab) specific for the proteins are useful for

CC detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body  
 CC fluids or tissue samples. HHV8 infections can be treated with antisense  
 CC or triplex forming molecules or agents that bind specifically to the  
 CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,  
 CC while the protein can be used in protective vaccines. Ab may also be used  
 CC to differentiate between lymphomas, and HHV8 may be implicated in many  
 CC other lymphoproliferative diseases such as lymphomas, leukaemia,  
 CC splenomegaly and mycosis fungoides. Cells and animals containing the  
 CC nucleic acid are useful for drug screening. HHV8-derived peptides can be  
 CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene  
 CC can be inhibited with methotrexate. These can also be used to determine  
 CC the immune status of a patient infected with HIV. HHV8 derived protein  
 CC viral MIP III may be used as an anti-inflammatory agent for,  
 CC e.g. treating rheumatoid arthritis. This sequence is stated as containing  
 CC 81 open reading frames. 32579 A; 37795 C; 35758 G; 31375 T;  
 SQ Sequence 137507 BP;

Query Match 6.3%; Score 16; DB 1; Length 137507;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 79 GGCAGCTTTTCTGTC 94  
 |||||||  
 Db 108877 GGCAGCTTTTCTGTC 108862

Search completed: August 5, 1999, 17:40:18  
 Job time: 6253 sec

**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:10:20 ; Search time 3264.21 Seconds

(without alignment cell updates/sec  
214.345 Million cell updates/sec)

Title: US-09-049-696-5

Perfect score: 220  
Sequence: 1 CTATAGTGAATCTGTCTACA.....GACCAATCTGGAAGCATGCGC 220

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_st: \*  
14: gb\_sts: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_vl: \*  
18: em\_fun: \*  
19: em\_hlg: \*  
20: em\_hum1: \*  
21: em\_hum2: \*  
22: em\_in: \*  
23: em\_om: \*  
24: em\_of: \*  
25: em\_ov: \*  
26: em\_pat: \*  
27: em\_ph: \*  
28: em\_pl: \*  
29: em\_ro: \*  
30: em\_sts: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_vl: \*  
34: gb\_hlg1: \*  
35: gb\_hlg2: \*  
36: gb\_in1: \*  
37: gb\_in2: \*  
38: em\_ba1: \*  
39: em\_ba2: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	220	100.0	3311	11	AF039400	Homo sapi

2	220	100.0	2826	42	AF127036	AF127036 Homo sapi
3	218	99.1	35278	11	AF039401	AF039401 Homo sapi
4	162	73.6	2937	12	AB017156	AB017156 Bos taurus
5	103.2	46.9	3317	3	AF001261	AF001261 Bos taurus
6	103.2	46.9	3288	3	AF001262	AF001262 Bos taurus
7	103.2	46.9	2820	3	AF001263	AF001263 Bos taurus
8	103.2	46.9	1265	3	AF001264	AF001264 Bos taurus
9	99.4	45.2	3415	42	AF043976	AF043976 Homo sapi
10	99	45.0	3022	12	AF047838	AF047838 Mus muscu
11	99	45.0	3471	12	AF052746	AF052746 Mus muscu
12	96.8	44.0	2984	3	BF036445	BF036445 Bos taurus
13	82.4	37.5	3604	9	AB026833	AB026833 Homo sapi
14	38.2	17.4	122459	35	AC006944	AC006944 Mus muscu
15	37	16.8	191857	34	CEY32B12	CEY32B12 Homo sapi
16	37	16.8	189165	35	AC007237	AC007237 Homo sapi
17	37	16.8	40662	36	CEY32B1B	CEY32B1B Homo sapi
18	36.2	16.5	133925	11	AC004562	AC004562 Caenorhab
19	36.2	16.5	123211	11	AC005208	AC005208 Homo sapi
20	35.8	16.3	219447	42	HUMFNG6PD	HUMFNG6PD Homo sapi
21	35.6	16.2	4855	37	AF062652	AF062652 Tetrahyme
22	35.4	16.1	11026	12	MUSWVDP	MUSWVDP Mus muscu
23	35.2	16.0	115564	34	HS1170D6	HS1170D6 Homo sapi
24	35.2	16.0	115564	9	HS1170D6	HS1170D6 Homo sapi
25	35.2	16.0	118869	42	AC004903	AC004903 Human DNA
26	35	15.9	134450	11	HUAC03049	HUAC03049 Homo sapi
27	35	15.9	142228	34	AC004123	AC004123 Homo sapi
28	35	15.9	168468	42	AC005510	AC005510 Homo sapi
29	34.8	15.8	39478	36	CEY53F1	CEY53F1 Caenorhabd
30	34.6	15.7	92797	11	AF038458	AF038458 Homo sapi
31	34.4	15.6	46462	36	CELC33F10	CELC33F10 Caenorhabd
32	34.2	15.5	79914	11	AC005159	AC005159 Homo sapi
33	34	15.5	75793	9	HS321F15	HS321F15 Human DNA
34	34	15.5	47003	11	AC005335	AC005335 Homo sapi
35	34	15.5	89382	34	CEY102A5_3	CEY102A5_3 Continuation (4 of
36	34	15.5	37994	36	CEFA0D4	CEFA0D4 Caenorhabd
37	33.8	15.4	40690	7	SCYLRO61W	SCYLRO61W X94607 S. cerevisia
38	33.8	15.4	1390	7	SCYLRO61W	SCYLRO61W X94607 S. cerevisia
39	33.8	15.4	324816	11	AC002368	AC002368 Homo sapi
40	33.8	15.4	150147	11	AC002543	AC002543 Homo sapi
41	33.6	15.3	148984	9	HS105D16	HS105D16 Human DNA
42	33.6	15.3	133120	9	HS393B23	HS393B23 Human DNA
43	33.6	15.3	149731	34	HS162013	HS162013 Homo sapi
44	33.6	15.3	192151	35	AC007282	AC007282 Homo sapi
45	33.4	15.2	135521	11	AC005858	AC005858 Homo sapi

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS	AF039400	3311 bp	mrna	PRI	15-DEC-1998	
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hcc1c1) mRNA, complete cds.					
ACCESSION	AF039400					
NTD	94009457					
VERSION	AF039400.1	GI:4009457				
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
AUTHORS	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.					
TITLE	Genomic cloning, molecular characterization, and functional analysis of human hcc1c1, the first human member of the family of Ca <sup>2+</sup> -activated Cl <sup>-</sup> channel proteins					
JOURNAL	Genomics 54 (2), 200-214 (1998)					
MEDLINE	99047526					
REFERENCE	2 (bases 1 to 3311)					
AUTHORS	Gruber,A.D., Elble,R. and Pauli,B.U.					
TITLE	Direct Submission					

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers  
source

1. 3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. 3311  
/gene="hCLCA1"  
352. 3096  
/gene="hCLCA1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPKSSVFILILHLEGLSLSLIQLNNNGYEGIVVAIDPNVP  
EDTLIQIKDMVTOASLYFEATGKREFYFNVAIILPETWKTADYVPRPLETKYNA  
DVLVAESTPSPNDPEPTPEOMGNGCEGERIHLTPDIAKKAIEYPOGKAIEYEMNA  
LRMGVDEYNDEKFEYLSNGRIQAVRCGAGITGNVYKCGGSCYTRCKGFENVTGL  
YKGCERFVLOSRTERKASIMFAOHVDSIVERCTEONHNKEAPNOKONCRSWETV  
RDSDFEKKTPMTQPPNPTSLQIGRIYCLVLDSSGMATNRLNLNQGLFL  
LQVLELSWGAVTFDSAAHVOSELQINSQSDKDTLAKRLPAASGGSISGELRSA  
FTVIRKRYPIDGSEIVLTDGEDNTISGCFNEVKOSGAIHTVALGPSAOLEELSKA  
MTGLOIYASDOVQNNGLIDAFGLSSNGAVSORSIOLESKGLTLNOSOMNGTVY  
DSTVGLDTLELITWTOPOIILMDPSGOGKFVVDKNTKMAIOLIPGIAKVTMY  
SLOASOTLITVTSRASNAITLPIVTSKNTKTSKPSPLVYVYANIRGASPIILRA  
SVTALLESVNGKTVTELDNGAGADATKDDGYTSRYFTTYDNGRISVYRALGVN  
AARRRVIPOOGSALYIPGMINDEIQWNPPELINDDVQHKQVCFSTRSSGGSFVAS  
DVPNAPIPLDFPQGITDKAEIHGSLINLTWTPAGDDYDHGTAHXYIIRISTSLD  
LRDFNESLQVNTTALIPKEANSEVLEFPENITFENGDTLFLAIQAVDKVLSKEI  
SNIRAVSLFIPQTPPETPSPDETSAFCPNHINSTIPGHIILKIMMKWIGELQLSIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 220; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 8e-49;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTATAGTGAATTCCTGTACAGAACAAAACACACAAAGAGCTCCAAACAGCAAAATC 60  
DB 1085 CTATAGTGAATTCCTGTACAGAACAAAACACACAAAGAGCTCCAAACAGCAAAATC 1144  
QY 61 AAAAATGCAATCTCCGAGACATGGGAAGTATCGTGATCTGAGACCTTTAAGAAA 120  
DB 1145 AAAAATGCAATCTCCGAGACATGGGAAGTATCGTGATCTGAGACCTTTAAGAAA 1204  
QY 121 CCACTCCTATGACACACAGCCACCAATCCACCTTCTCATTTGCTGAGATTGACAAA 180  
DB 1205 CCACTCCTATGACACACAGCCACCAATCCACCTTCTCATTTGCTGAGATTGACAAA 1264  
QY 181 GAATGTGTGTTAGTCTTGTGACAAATCTGGAAGCATGGC 220  
DB 1265 GAATGTGTGTTAGTCTTGTGACAAATCTGGAAGCATGGC 1304

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CaCC1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
AUTHORS 1 (bases 1 to 2826)  
TITLE Agnel,M., Vernet,T. and Culouscou,J.-M.  
Channel Cloning of three human homologs of bovine epithelial chloride channel

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel,M. and Culouscou,J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Neuilly-Malmaison 92500, France  
FEATURES Location/Qualifiers  
source

1. 2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. 2826  
/gene="CaCC1"  
5. 2749  
/gene="CaCC1"  
/note="CaCC1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="A025487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPKSSVFILILHLEGLSLSLIQLNNNGYEGIVVAIDPNVP  
EDTLIQIKDMVTOASLYFEATGKREFYFNVAIILPETWKTADYVPRPLETKYNA  
DVLVAESTPSPNDPEPTPEOMGNGCEGERIHLTPDIAKKAIEYPOGKAIEYEMNA  
LRMGVDEYNDEKFEYLSNGRIQAVRCGAGITGNVYKCGGSCYTRCKGFENVTGL  
YKGCERFVLOSRTERKASIMFAOHVDSIVERCTEONHNKEAPNOKONCRSWETV  
RDSDFEKKTPMTQPPNPTSLQIGRIYCLVLDSSGMATNRLNLNQGLFL  
LQVLELSWGAVTFDSAAHVOSELQINSQSDKDTLAKRLPAASGGSISGELRSA  
FTVIRKRYPIDGSEIVLTDGEDNTISGCFNEVKOSGAIHTVALGPSAOLEELSKA  
MTGLOIYASDOVQNNGLIDAFGLSSNGAVSORSIOLESKGLTLNOSOMNGTVY  
DSTVGLDTLELITWTOPOIILMDPSGOGKFVVDKNTKMAIOLIPGIAKVTMY  
SLOASOTLITVTSRASNAITLPIVTSKNTKTSKPSPLVYVYANIRGASPIILRA  
SVTALLESVNGKTVTELDNGAGADATKDDGYTSRYFTTYDNGRISVYRALGVN  
AARRRVIPOOGSALYIPGMINDEIQWNPPELINDDVQHKQVCFSTRSSGGSFVAS  
DVPNAPIPLDFPQGITDKAEIHGSLINLTWTPAGDDYDHGTAHXYIIRISTSLD  
LRDFNESLQVNTTALIPKEANSEVLEFPENITFENGDTLFLAIQAVDKVLSKEI  
SNIRAVSLFIPQTPPETPSPDETSAFCPNHINSTIPGHIILKIMMKWIGELQLSIA  
"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 220; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 8e-49;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTATAGTGAATTCCTGTACAGAACAAAACACAAAGAGCTCCAAACAGCAAAATC 60  
DB 738 CTATAGTGAATTCCTGTACAGAACAAAACACAAAGAGCTCCAAACAGCAAAATC 797  
QY 61 AAAAATGCAATCTCCGAGACATGGGAAGTATCGTGATCTGAGACCTTTAAGAAA 120  
DB 798 AAAAATGCAATCTCCGAGACATGGGAAGTATCGTGATCTGAGACCTTTAAGAAA 857  
QY 121 CCACTCCTATGACACACAGCCACCAATCCACCTTCTCATTTGCTGAGATTGACAAA 180  
DB 858 CCACTCCTATGACACACAGCCACCAATCCACCTTCTCATTTGCTGAGATTGACAAA 917  
QY 181 GAATGTGTGTTAGTCTTGTGACAAATCTGGAAGCATGGC 220  
DB 918 GAATGTGTGTTAGTCTTGTGACAAATCTGGAAGCATGGC 957

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
DEFINITION complete cds.

ACCESSION AF039401  
NTD 94009459  
KEYWORDS AF039401.1 GI:4009459  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source 1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..24522..24628,26644..26859,27447..27708,28766..31840..32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="GI:4009460"  
/translation="MGPEKSVFLLIHLLEGALNSLTQLNNGGIVAVIDPVP  
EDETLQOKDVTQASLYLFRATGRFRFFKRVALLIPETWTKADYAPKLETAKNA  
DLVAVESTIPGNDPEPTQWNGCNGEKGKIHLPDPFIAKRLAETGPOGKAVVHEAH  
LWGVGEYVNDNEKFLYSNGRIQAVKCSAGITGVNVKKGCGSCYTKCNKVTGL  
YKGECEYVQSROTERKASIMFAQVDSIYEECTEDHNEAKNKQKCNLSTWEVI  
RSEDFEKTPTPTQPPNPTFSLIOIGRIYCLVLDKSGSMGRNLRLNAGOLEL  
LOTVELSGVAVTFDSAAHVSSELIOISGSDPTLAKRLPAASGGSISGSRSA  
FTVIRKKYPTDSEIVLLDGEDNTISGCFNEKSGCALIHTVALGSPAAQLEELK  
MTGGLQITASDOVQNGGLIDAFGLSSGGAQSRIQLESGLTLQNSQWNGVTIV  
DSVAGKDTFLITWTQPPQILLMDPSGKGOGKGFVVDKNTKAYLQIPGIARVGTWY  
SIOASSQTLITVTSRASNAIPLPIVTSKTKNDTSKPSPLVYVAMINQAGSPILRA  
SVTALIESVNGKIVTLELDNGAGADATDDQVSYRFTYDNGRYSKVVALGGVN  
AARRVYPOQSGALVPGWIEQNDIOMNPPREINKDDQVHGVCESSRSGSGSTVAS  
DVPNAPIPDLFPFGQITDLNAELHSGSLNLTWTAAGDIDYDGTAKHTIIRISTSLD

1LDKRNESIQVNTALIPKRNSEEVLEFKPENITFENGCTDLFLAQAQVYDKSEI  
SNIAVSLFIPQIPETPSPDERSAPCPNHINSTITGIIHLKIMMNIGSLQLSIA  
"  
6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 99.1%; Score 218; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 2.4e-48;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 ATAGTGAATTCGTGTGACAGACAAACCAACCAAGAGCTCCAAACAGCAAAATCAA 62  
DB 18591 ATAGTGAATTCGTGTGACAGACAAACCAACCAAGAGCTCCAAACAGCAAAATCAA 18650  
QY 63 AAATGCAATCTCGAAGCACATGGGAGTGAATCCGTGATTCGAGGACTTTAAGAAAAC 122  
DB 18651 AAATGCAATCTCGAAGCACATGGGAGTGAATCCGTGATTCGAGGACTTTAAGAAAAC 18710  
QY 123 ACTCCATGACAAACACAGCAGCACAATCCACCTTCTCATTTGCTGCAGATTGACAAAGA 182  
DB 18711 ACTCCATGACAAACACAGCAGCACAATCCACCTTCTCATTTGCTGCAGATTGACAAAGA 18770  
QY 183 ATGTGTGTTAGTCTTGACAAATCTGGAAGCATGGC 220  
DB 18771 ATGTGTGTTAGTCTTGACAAATCTGGAAGCATGGC 18808  
RESULT 4  
AB017156 AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS Mus musculus gob-5 mRNA, complete cds.  
DEFINITION AB017156  
ACCESSION 93721911  
NTD

```

VERSION      AB017156.1 GI:3721911
SOURCE       Mus musculus adult intestine goblet cell cDNA to mRNA.
ORGANISM     Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
TITLE        1 (sites)
JOURNAL      Komiyama, Y. and Hirohashi, S.
MEDLINE      Cloning and identification of the gene gob-5, which is expressed in
AUTHORS      intestinal goblet cells in mice
JOURNAL      Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)
REFERENCE    2 (bases 1 to 2937)
AUTHORS      Komiyama, Y. and Hirohashi, S.
JOURNAL      Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru
TITLE        Tokoda, Tsukuba, Idakaki 300-2635, Japan
JOURNAL      (E-mail: tkomihncp.jst.go.jp, Tel: 81-298-47-7563,
              Fax: 81-298-47-5226)
FEATURES     Location=Qualifiers
              source
              1..2937
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /cell_type="goblet cell"
                /dev_stage="adult"
                /tissue_type="intestine"
                15..2756
                /gene="gob-5"
                15..2756
                /gene="gob-5"
                /codon_start=1
                /protein_id="BA03743.1"
                /db_xref="PID:d1034712"
                /db_xref="PID:g3721912"
                /db_xref="PID:g3721912"
                /translation="MESLKSPPVLLIILHLEGVSESLIOLNNNGEYVIAIDHDVY
                EDLILQIKRMVMOASPYLFEAGKRRFFNNVALLIPESKAREPIRPLREFKNA
                DVLVSTSPLENDEPIETHITACGEGKGRILHIDPFLGKLLIOTGPODRTFVEMAH
                FRWCVFENYNDENKFFYLSKGRQVAVRCSAALTGNQVRCOGGSCITNGKCIDRVYGG
                LYKNQCVFDPHONERKASIMFNQINSIVVECTEKNNQDAPODNGRCMLRSTEWEL
                IOESDEKQTPMTAQPAPFPELSLIQIORTVCYLVDLRSGLMDDRLRNQASRLLE
                LLOVEQSGVNGVTFDESAAYVOSLEKQNSGADBDLLIKLPTVSAGSTIGSLR
                AFVYIKKKYPTDGSSEIVLITDGEDNTSLSCSDIKVQSGAIIHTVALGPAALKTEQLS
                KMTGGLQITSSDQVQNNGLVVAFAALBSGNAAIKQHSIQLESRYVQNNQMNNGSVY
                VDSVSGKDTLELITWTHPTLITMDPSGVNGEFLDTTIVKVALQVPGTAKGFWM
                YSIQASSOTLTLIVTSRAASATLPEPIVTVPVNKNTEGFPSPVTVYASIRQASPIILR
                ASVALLESVNGKTVTLELDNGAGADATKNDGYVSRFFAFDANGRSYVIMLGVN
                LVSQRAAPKPNRMYIDGWEIDEGEVNRPPEPYSYODKOLCFRSSEGSSEVAT
                VPAAPIPDLEPFCOITDKASIOGQNVNLTWTAAPGDYDHGASNLTIMRSIYDV
                LRHNETSLQVNTGTGLIPKESASEEIEFELGMPFGNGTIFALQAVXKSNLSE
                SNLRVSYFIADQEPPIEDSTPPCPDISINSTIPGILHVLKIMKMLGEMOVTGLH
BASE COUNT  860 a 718 c 693 g 666 t
ORIGIN
Query Match      73.6%; Score 162; DB 12; Length 2937;
Best Local Similarity 83.9%; Pred. No. 1.6e-33;
Matches 183; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 1 CTATAGTTGAATTCGTGTACAGAACAAACACACACAAAGAAAGCTCCAAACAAACAAATC 60
Db 751 CTGTGGTTGAATTCGTGTACAGAAAATAATACACATCAAGAGCCCAATATGACCAAAAC 810
QY 61 AAAAATGCAATCTCCGAGACACATGGGAAGTATCCGTGATCTCTAGAGACTTTAAGAAA 120
Db 811 AAGCATGGAATCTCCGAGACACAGTGGGAAGTATCCACAGAAATCTGAGACTTCAAGCAA 870
QY 121 CCACATCCATATACACACAGCCACCAAAATCCCACTTCTCATTTGGAGATTTGACAAA 180
Db 871 CCACATCCCATATACACAGCCACCAAGCCACTCTGACACCCACTTCTCACTGCTGCAAAATG 930
QY 181 GAATTGTGTGTAGTCTCTTGAACCAATCTGGAAGCATG 218

```

DB	931	GAATGTGTGCTTACTTCTTGTATTAAGTCCGGAGCATG	968
RESULT	5		
AF001261			
LOCUS			
DEFINITION			
ACCESSION			
NID			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
SOURCE			
CDS			
BASE COUNT			
ORIGIN			
Query Match	46.9%	Score 103.2;	DB 3; Length 3317;
Best Local Similarity	69.5%	Pred. No. 5.3e-18;	
Matches 157;	Conservative	0; Mismatches 63;	Indels 6; Gaps 1.
1	CTATAGTTGATTCCTGTGACAGAAACCAACCAAGCAAGCTCCAAACAAACCAATTC	60	
799	CTGTGACATGCAATTTTGTTCACGAAAAAACAACACACATACACGAAGCTCCAAACCTACAAACCA	858	
61	AAAAATGCAATCTCCGAGACACATGGGAAGTATCCGTGATTTCTGAGGACTTTAAGAAAA	120	

```

Db      859 AATGTGATGCAAGCAACACCTGGATGATCATGAACTGTGTGACTTTCAGAAATA 918
      121 CCACCTCTATGACA-----ACACAGCCACCAAAATCCCACTTCATGTCGTGAGATTG 174
      Db      919 CATCTCCCATGACAGAAATGAATCCACGACATCTCTATCATTTTATGTCCTCAATGCA 978
      Qy      175 GACAAAGAAATGTGTAGTTCCTTGACAAATCTGGAAGCATGCG 220
      Db      979 AACAGCGGGTAGTCTGTGTGACTTGATAAATCTGGAAGCATGTC 1024

RESULT 6
AF001262 3288 bp mRNA MAM 19-NOV-1997
LOCUS Bos taurus clone 2 endothelial adhesion molecule Lu-ECAM-1 mRNA,
DEFINITION complete cds.
ACCESSION AF001262
VERSION 92623764
KEYWORDS AF001262.1 GI:2623764
SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 3288)
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,
Goodwin, A. and Pauli, B. U.
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an
JOURNAL endothelial chloride channel
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 3288)
Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,
Goodwin, A. and Pauli, B. U.
TITLE Direct Submission
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., WMC C4-146,
Ithaca, NY 14853, USA
FEATURES
source Location/Qualifiers
1..3288
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_type="aortic endothelial cells"
/clone="2"
/tissue_type="lung"
59..2443
/note="endothelial adhesion molecule; chloride channel"
/codon_start=1
/product="Lu-ECAM-1"
/protein_id="AAB8530.1"
/db_xref="PID:92623765"
/db_xref="GI:2623765"
/translation="MVLCLNVLFLTLHLPGKSSMVLINNGDGIIVAINPSVE
DEKLEIKENKWTASTYLPHATKRRVFRVNSILIPMTKSKSEYFIPKDSYQAD
VIVANPYLKTGDDPTLDYGRGEGKATIHPTNPLTNHHTIGSRGVVHEAHL
RMGIFDEVNVOPEYISRNKITEARCSHTHTGIVNFKCPGSCITSLCRDSOTG
LYEAKCTFLPKRSQAKESIMFSLHSVECTEKTNTTEAPNLONMCKGKSTWDY
INNSYDFONTSPMTENPPTHTFSLSKORVCLVLDKSGMSAEDRLPOMQAAE
LYLLOVIEKSGLVGWTDFVSAEIONHLRTIDNVYOKITAKLPQVANGGSIORGL
KAGDAIHSOSQSGSEIILITGDEMEINSCFEDYKRSALITHTILGSAKELE
TLNMTGGRFRFANKDITGLTNARSRISRSRGSITQOAIQLESRAKLTGKRNVGIV
PVDSTVGNDTFVVTWIOKPEIVLQDPKRGKRTSDREKDLNRSARLQIPGIAET
GWTYSLNHNHASSQMLVTYVTRARSPITIPVIAAHMSQHTAHYPSPMIVYQVSO
GFLPYLGISVIAIETEDGHQVTLLEMDNGAGRDYVKNKGISRSFTDYGNGRSLK
VHAORNNTRNLNROPONKVLVYGVYENGKIIINPRPEVKDILAAKIDFESRLT
SGGSFTVGAPPRGNHPSVPFPRSKITDLEAKFKEDYIQLSWTAPONVLDKKGSSP
BASE COUNT 1043 a 707 c 645 g 893 t
ORIGIN
      46.9% Score 103.2; DB 3; Length 3288;
      Best Local Similarity 69.5%; Pred. No. 5.3e-18;
      Matches 157; Conservative 0; Mismatches 63; Indels 6; Gaps 1;
```

```

Qy      1 CTAATGTAATTCGTGTACAGAACAAACCACACAGAGAGCTCCAAACGAAATC 60
      Db      795 CTTGTGACTGAATTTGTATACAGAAAAACACACAATACAGAGCTCCAAACGAAACAA 854
      Qy      61 AAAAATGCAATCTCGAAGACACATGGGAAGTATCCGATATCTAGAGCTTAAAGAA 120
      Db      855 AATGTGCAATGCAAGCAACACATGGAGATGATGATGAACTGTGTGACTTTCAGAAATA 914
      Qy      121 CCACCTCTATGACA-----ACACAGCCACCAAAATCCCACTTCATGTCGTGAGATTG 174
      Db      915 CATCTCCCATGACAGAAATGAATCCACGACATCTCTATCATTTTATGTCCTCAATGCA 974
      Qy      175 GACAAAGAAATGTGTAGTTCCTTGACAAATCTGGAAGCATGCG 220
      Db      975 AACAGCGGGTAGTCTGTGTGACTTGATAAATCTGGAAGCATGTC 1020

RESULT 7
AF001263 2820 bp mRNA MAM 19-NOV-1997
LOCUS Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA,
DEFINITION complete cds.
ACCESSION AF001263
VERSION 92623766
KEYWORDS AF001263.1 GI:2623766
SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 2820)
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,
Goodwin, A. and Pauli, B. U.
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an
JOURNAL endothelial chloride channel
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 2820)
Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,
Goodwin, A. and Pauli, B. U.
TITLE Direct Submission
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., WMC C4-146,
Ithaca, NY 14853, USA
FEATURES
source Location/Qualifiers
1..2820
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="3"
/tissue_type="lung"
/cell_type="aortic endothelial cells"
194..2656
/note="endothelial adhesion molecule; chloride channel"
/codon_start=1
/product="Lu-ECAM-1"
/protein_id="AAB8531.1"
/db_xref="PID:92623767"
/db_xref="GI:2623767"
/translation="MVLCLNVLFLTLHLPGKSSMVLINNGDGIIVAINPSVE
DEKLEIKENKWTASTYLPHATKRRVFRVNSILIPMTKSKSEYFIPKDSYQAD
VIVANPYLKTGDDPTLDYGRGEGKATIHPTNPLTNHHTIGSRGVVHEAHL
RMGIFDEVNVOPEYISRNKITEARCSHTHTGIVNFKCPGSCITSLCRDSOTG
LYEAKCTFLPKRSQAKESIMFSLHSVECTEKTNTTEAPNLONMCKGKSTWDY
INNSYDFONTSPMTENPPTHTFSLSKORVCLVLDKSGMSAEDRLPOMQAAE
LYLLOVIEKSGLVGWTDFVSAEIONHLRTIDNVYOKITAKLPQVANGGSIORGL
KAGDAIHSOSQSGSEIILITGDEMEINSCFEDYKRSALITHTILGSAKELE
TLNMTGGRFRFANKDITGLTNARSRISRSRGSITQOAIQLESRAKLTGKRNVGIV
PVDSTVGNDTFVVTWIOKPEIVLQDPKRGKRTSDREKDLNRSARLQIPGIAET
GWTYSLNHNHASSQMLVTYVTRARSPITIPVIAAHMSQHTAHYPSPMIVYQVSO
GFLPYLGISVIAIETEDGHQVTLLEMDNGAGRDYVKNKGISRSFTDYGNGRSLK
VHAORNNTRNLNROPONKVLVYGVYENGKIIINPRPEVKDILAAKIDFESRLT
SGGSFTVGAPPRGNHPSVPFPRSKITDLEAKFKEDYIQLSWTAPONVLDKKGAGSP
MSRSHQVAKVLEIQLQHQSFQ"
```



BASE COUNT	923 a	710 c	661 g	728 t
ORIGIN				
Query Match	45.0%	Score 99;	DB 12;	Length 3022:
Best Local Similarity	68.4%;	Pred. No. 6.9e-17;		
Matches 154; Conservative	0;	Mismatches 65;	Indels 6;	Gaps 1;
OY	1	CTATGTTGAATTTCGTCTGTACAGAAACAACCAACCAAGAAGCTCCCAACCAACGAATAATC	60	
Db	751	CTGGTGGTAATTTTGCCACAGAAAAAACCACAAAGCAGAAAGCCCCAACCTACAAAACA	810	
OY	61	AAAAATGCATCTCCGAGACACATGGGAAGATGCCGTGATCTGTAGAGCATTTAAGAAA	120	
Db	811	AAATGTGCATCTCGAAGAACAGCTGGATGTATATCAACAGCTCTGACTTACGAATG	870	
OY	121	CCACTCTCATG-----ACAACACAGCCACCAAAATCCCACCTTCTCATTTGCTGCAGATTG	174	
Db	871	CCCCCTCCATGAGAGAACAGAGCCCTCCTCCACCTACATTTTCACTGCTCAAGTCCA	930	
OY	175	GACCAAAGATGTGTGTTTACTTCCTTAGCAAAATCTGGGAAGCATGG	219	
Db	931	GAAAGCGAGTGTGTGTTGTTGCTTGATTAATCTGGAAGCATGG	975	
RESULT 11				
LOCUS	AF052746	3471 bp	mRNA	ROD 08-Apr-1999
DEFINITION	Mus musculus chloride channel CaCC mRNA, complete cds.			
ACCESSION	AF052746			
NID	93560546			
VERSION	AF052746.1	GI:3560546		
KEYWORDS	house mouse.			
SOURCE	Mus musculus			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 3471)			
AUTHORS	Romfo,L., Musante,L., Cinti,R., Serl,M., Moran,O., Zagarra-Moran,O. and Galietta,L.J.V.			
TITLE	Characterization of a murine gene homologous to the bovine CaCC chloride channel			
JOURNAL	Gene 228 (1-2), 181-188 (1999)			
MEDLINE	99173882			
REFERENCE	2 (bases 1 to 3471)			
AUTHORS	Romfo,L., Musante,L., Cinti,R., Moran,O., Serl,M. and Galietta,L.J.V.			
TITLE	Direct Submission			
JOURNAL	Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto Giannina Gaslini, Genova I-16148, Italy			
FEATURES	Location/Qualifiers			
SOURCE	1..3471			
CDS	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/db_xref="dbEST:AA245677"			
	47..2752			
	/codon_start=1			
	/product="chloride channel CaCC"			
	/protein_id="AAC35003.1"			
	/db_xref="PID:g3560547"			
	/db_xref="GI:3560547"			
	/translation="VWPGDLVLFLTIHLIQLNTSSKAVHLSNSQIEGVVAIINPSPVEE DELLPISEIKEMVQAQSTLYECASQGRVFENISLIVPMTRSKREYLMPRESYDKAD VIYADHLIGDDPDYLOYGOCGDGRGYIHFTFNPLITLMRLITGPGRGFVHEMAHL RMCVPEYNVDOPFMSRKNTIETARSITGITGNVNANEERGNCVIRACRDSKTSL YEKCFEIPDKIOITAGASIMFMGNINSNVSECTEKNNAPARNONKNCRBSRWIVYI KRADRONAPRNMGTEAPPPTPSILKSRRRYCVLVDRSGSNDEKERLRIRMNDAAL VTQIVTESKWGLVTFDASAHIQNIYIKITSSSDYTITANLPQOASGETSIICHGLQ AGQVATESDOSTSGSEIVLITLTGEDNGISCFEAVSRAIIHTLTLGPSARELEET			







Mon Aug 9 13:26:19 1999

us-09-049-696-5\_1.rge

Page 10

\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

**FEATURES**  
**SOURCE**

```
Location/Qualifiers
1. 191857
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
```

	a	c	g	t	others
BASE COUNT ORIGIN	58119	33310	32923	58630	8875

Query Match	16.88;	Score 37;	DB 34;	Length 191857;
Best Local Similarity	57.38;	Pred. NO. 1.3;		
Matches 67; Conservative	0;	Mismatches 50;	Indels 0;	Gaps 0;

[illegible]

Search completed: August 5, 1999, 23:10:40  
Job time: 7523 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:36:19 ; Search time 3268.17 Seconds

(without alignments)  
246.198 Million cell updates/sec

Title: US-09-049-696-8

Perfect score: 253  
Sequence: 1 AACCAAGTGTGTCATC.....GGCAGCTATCTGTGACAG 253

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pl3.\*  
10: gb\_pl4.\*  
11: gb\_pl5.\*  
12: gb\_pl6.\*  
13: gb\_pl7.\*  
14: gb\_pl8.\*  
15: gb\_pl9.\*  
16: gb\_pl10.\*  
17: gb\_pl11.\*  
18: gb\_pl12.\*  
19: gb\_pl13.\*  
20: gb\_pl14.\*  
21: gb\_pl15.\*  
22: gb\_pl16.\*  
23: gb\_pl17.\*  
24: gb\_pl18.\*  
25: gb\_pl19.\*  
26: gb\_pl20.\*  
27: gb\_pl21.\*  
28: gb\_pl22.\*  
29: gb\_pl23.\*  
30: gb\_pl24.\*  
31: gb\_pl25.\*  
32: gb\_pl26.\*  
33: gb\_pl27.\*  
34: gb\_pl28.\*  
35: gb\_pl29.\*  
36: gb\_pl30.\*  
37: gb\_pl31.\*  
38: gb\_pl32.\*  
39: gb\_pl33.\*  
40: gb\_pl34.\*  
41: gb\_pl35.\*  
42: gb\_pl36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	253	100.0	3311	11	AF039400	Homo sapi

253	100.0	2826	42	AF127036	AF127036 Homo sapi	
109	43.1	35278	11	AF039401	AF039401 Homo sapi	
4	22	8.7	2937	12	AB017156	AB017156 Mus muscu
5	18	7.5	169526	35	AC007081	AC007081 Drosophila
6	18	7.1	113143	9	HS544A11	AL023281 Human DNA
7	18	7.1	136222	11	AC004099	AC004099 Homo sapi
8	18	7.1	1386	12	RN36785	U36785 Rattus norv
9	18	7.1	3804	12	RRASGRF	X67241 R. rattus mr
10	18	7.1	190000	35	AC006427	AC006427 Homo sapi
11	18	7.1	56479	42	AC006034	AC006034 Homo sapi
12	17	6.7	208430	1	BSUB0005	Z99108 Bacillus su
13	17	6.7	9444	1	D78508	D64000 Synchocyst
14	17	6.7	146271	1	SYCLSRB	U14855 Bos taurus
15	17	6.7	6349	3	BOVNEURX1A	U27351 Fells catus
16	17	6.7	2534	3	FCU27351	X17255 Bacterioph
17	17	6.7	19680	6	POT3111G	Z98978 S. pombe chr
18	17	6.7	34595	7	SPAC27E2	AC004411 Arabidops
19	17	6.7	106329	8	ATAC004411	U63550 Fragaria x
20	17	6.7	1408	8	FXU63550	AB011150 Homo sapi
21	17	6.7	5703	9	AB011150	D87452 Human mRNA
22	17	6.7	4461	9	D87452	D83850 Human DNA s
23	17	6.7	132981	9	HS82J11	D26350 Human mRNA
24	17	6.7	10524	9	HUMH21	AL023581 Human DNA
25	17	6.7	137947	10	HS4514	AL031672 Human DNA
26	17	6.7	100958	10	HS691N24	AC003090 Homo sapi
27	17	6.7	94255	11	AC003090	AC004499 Homo sapi
28	17	6.7	100452	11	AC004499	AC004601 Human 11p
29	17	6.7	132867	11	AC004601	AC004774 Homo sapi
30	17	6.7	161652	11	AC004774	AC005684 Homo sapi
31	17	6.7	143865	11	AC005684	AC005685 Homo sapi
32	17	6.7	173717	11	AC005685	AC005747 Homo sapi
33	17	6.7	166701	11	AC005747	AC006080 Homo sapi
34	17	6.7	191111	11	AC006080	AF001871 Mus muscu
35	17	6.7	1470	12	AF001871	AF047838 Mus muscu
36	17	6.7	3022	12	AF047838	AF052746 Mus muscu
37	17	6.7	3471	12	AF052746	AF050157 Mus muscu
38	17	6.7	245439	12	NRAC135G15	M96374 Rattus norv
39	17	6.7	5609	12	RATN1A	S76673 mTEF3 (mice
40	17	6.7	2469	12	S76673	AC005949 Homo sapi
41	17	6.7	155093	35	AC005949	AC005949 Homo sapi
42	17	6.7	145449	35	AC005949	AC007322 Homo sapi
43	17	6.7	168282	35	AC007322	AC007323 Homo sapi
44	17	6.7	262196	35	AC007323	AC007778 Homo sapi
45	17	6.7	167713	35	AC007778	

## ALIGNMENTS

RESULT 1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS	AF039400				
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA, complete cds.				
ACCESSION	AF039400				
NID	94009457				
VERSION	AF039400.1	GI:4009457			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	Gruber,A.D., Eibler,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Paul,B.U.				
TITLE	Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins				
JOURNAL	Genomics 54 (2), 200-214 (1998)				
MEDLINE	99047526				
REFERENCE	2 (bases 1 to 3311)				
AUTHORS	Gruber A.D., Eibler, R. and Paul, B. U.				
TITLE	Direct Submission				

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES Location/Qualifiers

source

1.3311

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="1p22-p31"

/tissue\_type="small intestine"

1.3311

/gene="hclcal"

352.3096

/gene="hclcal"

/note="transmembrane glycoprotein"

/codon\_start=1

/product="calcium-dependent chloride channel-1"

/protein\_id="AAC95428.1"

/db\_xref="pid:9409458"

/db\_xref="gi:409458"

/translation="MGPKSSVFLLILHLLEALSNSLIQLNNNGEGIVAIIDPNVP  
EDETLIQIKMVTQASLYLEAATGKREFKNVAILPETWTKADYAPKLETKMA  
DVALESTPGNDEPYEBQMGCEKGERIHLPTFIAGKLAEGPOGRAFEVEMAH  
LKGVEDEYNDEKFTYLSNGRIQAVRCASGIGTINVKKCGGSCYTRKCTFNKTVGL  
YKGEFEVLSROTEKASIMFAOHVDSIVERCTEONHNKEAPNOKNCONLSTREVI  
RDSDFPKTTPMTOPPNPTFSLIOIGQIVLYLVLDKSSMATGRNLNLAGOLEL  
LQTVELGSGWGVTFDSAAHVOSELIOINSQSDKDTLAKRLPAASGGTISGRLSA  
FTVIRKRYPTDSEIVLITDGEDNTISGFENVKOSGAIHTVALGPSAOGLELSK  
MTGLOTVASDOVONNGILDAFGALSSGNGAVSORSIOLESKGLLONSONMNGTVIV  
DSTVGKDTLELITWTOPPOILLMDPSGQKGFVVDKNTKMAVLOIPDAVGVIMKY  
SLOASGOTLITVTSRASNAITLPITVTSKTKNDKSKPSPLVYVANTROGASPLIRA  
SYTALIESNKGTVLELIDNGAGADATDDQVYSRYFTTYTNGRISRYKVALGYN  
AARRVYIQOSGALYIPGWIENDEITOMNPREINKDVOHQVCFSESTSGSFVAS  
DVPNAPIDLPFGQITDLKAEIHGSLINLTWAPGDYDGHAKYIIRISTILD  
LNDKFNESLQVNTALIPKEANSEVEFLFKPENITFENGTDLFIQAVDVKLSEI  
SNIRVSLFIPQTPPEPSPDETSAFCPNHINSTIPGIIHLIKIMMKWIGELQSLIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 253; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 2.1e-132;  
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACAAAGTGTGGCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAG 60  
DB 1628 AACAAAGTGTGGCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAG 1687  
QY 61 AGGAGCTGTCCAAAATGACAGAGGTTTACAGCATATGCTTACATCAAGTTCAGAAACA 120  
DB 1688 AGGAGCTGTCCAAAATGACAGAGGTTTACAGCATATGCTTACATCAAGTTCAGAAACA 1747  
QY 121 ATGGCCTCATTTGATGCTTTGGGGCCCTTTCATCAGAGAAATGAGCTGTCTCAGCGCT 180  
DB 1748 ATGGCCTCATTTGATGCTTTGGGGCCCTTTCATCAGAGAAATGAGCTGTCTCAGCGCT 1807  
QY 181 CCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGAACACGCGATGATGATGACAGAG 240  
DB 1808 CCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGAACACGCGATGATGATGACAGAG 1867  
QY 241 TGATCGTGGACAG 253  
DB 1868 TGATCGTGGACAG 1880

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (Caccl1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID \*94585468  
VERSION AF127036.1 GI:4585468

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 2826)

Aguel,M., Vernat,F. and Culouscou,J.-M.

Cloning of three human homologs of bovine epithelial chloride

channel

Unpublished

2 (bases 1 to 2826)

Aguel,M. and Culouscou,J.-M.

Direct Submission

Submitted (09-FEB-1999) Groupe Genomique, SYNHELABO, 10 Rue des

Carrieres, Rueil-Malmaison 92500, France

Location/Qualifiers

1.2826

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/tissue\_type="small intestine; colon"

1.2826

/gene="Caccl1"

5.2749

/gene="Caccl1"

/note="bovine epithelial chloride channel homolog"

/codon\_start=1

/product="calcium-activated chloride channel protein 1"

/protein\_id="AAD25487.1"

/db\_xref="pid:94585469"

/db\_xref="gi:4585469"

/translation="MGPKSSVFLLILHLLEALSNSLIQLNNNGEGIVAIIDPNVP  
EDETLIQIKMVTQASLYLEAATGKREFKNVAILPETWTKADYAPKLETKMA  
DVALESTPGNDEPYEBQMGCEKGERIHLPTFIAGKLAEGPOGRAFEVEMAH  
LKGVEDEYNDEKFTYLSNGRIQAVRCASGIGTINVKKCGGSCYTRKCTFNKTVGL  
YKGEFEVLSROTEKASIMFAOHVDSIVERCTEONHNKEAPNOKNCONLSTREVI  
RDSDFPKTTPMTOPPNPTFSLIOIGQIVLYLVLDKSSMATGRNLNLAGOLEL  
LQTVELGSGWGVTFDSAAHVOSELIOINSQSDKDTLAKRLPAASGGTISGRLSA  
FTVIRKRYPTDSEIVLITDGEDNTISGFENVKOSGAIHTVALGPSAOGLELSK  
MTGLOTVASDOVONNGILDAFGALSSGNGAVSORSIOLESKGLLONSONMNGTVIV  
DSTVGKDTLELITWTOPPOILLMDPSGQKGFVVDKNTKMAVLOIPDAVGVIMKY  
SLOASGOTLITVTSRASNAITLPITVTSKTKNDKSKPSPLVYVANTROGASPLIRA  
SYTALIESNKGTVLELIDNGAGADATDDQVYSRYFTTYTNGRISRYKVALGYN  
AARRVYIQOSGALYIPGWIENDEITOMNPREINKDVOHQVCFSESTSGSFVAS  
DVPNAPIDLPFGQITDLKAEIHGSLINLTWAPGDYDGHAKYIIRISTILD  
LNDKFNESLQVNTALIPKEANSEVEFLFKPENITFENGTDLFIQAVDVKLSEI  
SNIRVSLFIPQTPPEPSPDETSAFCPNHINSTIPGIIHLIKIMMKWIGELQSLIA  
"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 253; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 2.1e-132;  
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACAAAGTGTGGCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAG 60  
DB 1281 AACAAAGTGTGGCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAG 1340  
QY 61 AGGAGCTGTCCAAAATGACAGAGGTTTACAGCATATGCTTACATCAAGTTCAGAAACA 120  
DB 1341 AGGAGCTGTCCAAAATGACAGAGGTTTACAGCATATGCTTACATCAAGTTCAGAAACA 1400  
QY 121 ATGGCCTCATTTGATGCTTTGGGGCCCTTTCATCAGAGAAATGAGCTGTCTCAGCGCT 180  
DB 1401 ATGGCCTCATTTGATGCTTTGGGGCCCTTTCATCAGAGAAATGAGCTGTCTCAGCGCT 1460  
QY 181 CCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGAACACGCGATGATGATGACAGAG 240  
DB 1461 CCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGAACACGCGATGATGATGACAGAG 1520  
QY 241 TGATCGTGGACAG 253  
DB 1520 TGATCGTGGACAG 1520

Db 1521 TGATCGTGACAG 1533

RESULT 3  
AF039401

LOCUS AF039401 35278 bp DNA PRI 15-DEC-1998  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.

ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459

KEYWORDS

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 35278)

REFERENCE  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schneur,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human CLCA1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

FEATURES  
source  
1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="L1 transposable element"  
1587..1596  
/rpt\_type="dispersed"  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696,9801,15454..15631,18591..18809,19783..20010,22253,22427,24522,24628,26644..26859,27447..27708,28766,28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1562..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253,22427,24522,24628,26644..26859,27447..27708,28766,31840..32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="GI:4009459"  
/db\_xref="PID:94009459"  
/translation="MGPKSSVFILILHLEGLALSLSLQLNNNGIEGIVAIIDPNV  
EBETLIQIKIDVTAQSLYLEATGRRTFKVNAIILPTWKADYVPRKLETYNA  
DYLVAESTPGNDPEYTEMNGEGERHILPDILAKKLAIEYPOGKAPEHENA  
LRMGVDEYNNDKEFLYSNGRIQAVRSAGITLTVYKKCGGSGCTKGTETKVTGL  
YKGCPEYLOSROTERTKASIMFGRQVDSIYEFCEQHNHNEAPKOKOKNTRTEWTI  
RSEDFKKTTPPTTOPNPFTSLQIGQRIVCILVDKSGSMATGNLRLNKGQLFL

LQTVELGSGVGVAVTDSAAHVOSSELIQINSQSDROTIAKRLPAASGCTISGCLRSA  
FTVIRKRYPTDSEIVLLTDEGDNITSGCFNEVKOSGALIHVVALGPSAQLEIELSK  
MTGGIQTAVSDVONNGLIIDAFGALSNGAVSORSIOLESGLTLQNSQMNNGVTVY  
DSVAGQDLFLITWTQPOILLMPDSGKOGGFVDDKNTKAYLQIPIGIAYGWTXY  
SLQASQDLTLITVTSRASNAATLPITVTSKTKNDTSKEPSPLVYVANIINGQSPILRA  
SVTALIESVNGTIVTIELDNGAGDATTDDCVYSKFTYTDNCRYSKRVALGAVN  
AARRVTPQSSGALIPGWIENDEIQWNPREFINDDVOHROVCSRTSSGSGFPVAS  
DVFNAPIPDLPFPQGTTDLMAEIHGSLINTWTAPGDDYDGTAKYIIRLSTSLD  
LHDKNESLQVNTALIPKEANSSEVLEKPEINIFENGTDLFLIAQAVDKVLSKEI  
SNARVSLFIPPTPEPTSPDETSAPCPNINHNSIIPGIIHLIKIMWKIGELQLSIA  
"

6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15

BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN

Query Match 43.1%; Score 109; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 3.6e-51;  
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 AGGAGGTTACAGACATATGCTTCAGATCAAGTTCAAGAACATGGCCTCATGATGTTT 139  
|||||  
DB 24520 AGAGAGTTTACAGACATATGCTTCAGATCAAGTTCAAGAACATGGCCTCATGATGTTT 24579  
|||||

QY 140 TGGGGCCCTTTCATCAGAAATGAGACTGTCTCTCAGCGCTCCATCCAG 188  
|||||  
DB 24580 TGGGGCCCTTTCATCAGAAATGAGACTGTCTCTCAGCGCTCCATCCAG 24628  
|||||

RESULT 4  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS AB017156  
DEFINITION Mus musculus gob-5 mRNA, complete cds.  
ACCESSION AB017156  
NID 93721911

```

VERSION      AB017156.1 GI:3721911
KEYWORDS     GDB-5.
SOURCE       Mus musculus adult intestine goblet cell cDNA to mRNA.
ORGANISM     Mus musculus
REFERENCE    Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
AUTHORS      Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
TITLE        1 (sites)
JOURNAL      Komiya, T., Tanigawa, Y. and Hirohashi, S.
MEDLINE      Cloning and identification of the gene gob-5, which is expressed in
AUTHORS      intestinal goblet cells in mice
REFERENCE    Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)
JOURNAL      99160866
TITLE        2 (bases 1 to 2937)
AUTHORS      Komiya, T., Tanigawa, Y. and Hirohashi, S.
REFERENCE    Direct Submission
JOURNAL      Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru
AUTHORS      Komiya, T., Tanigawa, Y., Hirohashi, S. Cell Configuration Project, 5-9-9,
REFERENCE    Tokodai, Tsukuba, Ibaraki 300-2635, Japan
JOURNAL      (E-mail: tkom@hccp.jst.go.jp, Tel: 81-298-47-7563,
REFERENCE    Fax: 81-298-47-5226)
FEATURES
SOURCE       Location/Qualifiers
1..2937
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="goblet cell"
/dev_stage="adult"
/tissue_type="intestine"
15..2736
/gene="gob-5"
15..2756
/gene="gob-5"
/codon_start=1
/db_xref="PID:41034712"
/db_xref="PID:93721912"
/db_xref="GI:3721912"
/translation="MESLKSPEVLLILHLLEGLVLESESLIQLNNNGEGIVTADHDV
EDALQIHKDVTQASPYLFATGKREFKFNVALIIPESWKAKEEYRPRLEFKNA
DVLVSTSPINDDEPTEHIGACGKGRIRHITPDPLAGKICITGCPDRPTVEHRAH
FMGVNENNDKERTYLSGKQAVRCATITGKQVRCQSGSITNGKCVITDVTG
LYKDCVVPDPHONKASIMFNQINSVECTEKNHQNQEPNQNOCNRMSTWEY
IQSEDFKQTPMTPAPFPLSLQIQRIYCLVLDKSGSLNDNRNKNQASRLT
LIQVEQSGWQVMTFDSAAVYQSEIKOLNSGADRLIKHLPTVSAGSTICSGLT
AFTVKKRPEDGSEIVLLDGEDNTISSCFDLVQSGAIIHTVVLGAAAKKEIOLS
KMTGLQTYSSDOVQNGVDAFAALSGMAIAQHSIOLSRGVNLONNMNGSVI
VDSVSKDTLELITHTPTPTFTFIDPSCVQNGILDTTTRKATLOVPGTAKGVFK
YSIQASSQTLITVTSRAASATLPIPTVYVKNKTKGFPSPVTVYASIRGASPIIR
ASVTALESVNGKTVTLLELDNGAGADATKNDGVSRFTAFDANGARSVIMLGVY
TSDRQAPPKRPMYIDGWIEDGEVNRNPPETSYQDOLCFSRSSGSEVATN
VPAAPIDLPFCOITDLKASIOGNLVNLTWAPGDYDHGRASNTIIMSTSIYD
LSDHNTSLQVNTGLIPKASSEIEFEELGNGTNGTDFIALQAVDSKNSKET
SNIAVSFVIPAQEPRIEDSPRPDISINSTIGIHVLAIMKNMLEMAYTTGLH"
BASE COUNT  860 a 718 c 693 g 666 t
ORIGIN
Query Match      8.7% Score 22: DB 12: Length 2937:
Best Local Similarity 100.0%: Pred. No. 0.049:
Matches 22: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 64 AGCTGTCACAAATGACAGAGG 85
| | | | | | | | | | | | | | |
DB 1357 AGCTGTCCAAATGACAGAGG 1378

```

```

VERSION      AC007081.5 GI:4934937
KEYWORDS     HTG: HTGS_PHASE1.
SOURCE       fruit fly
ORGANISM     Drosophila melanogaster
REFERENCE    Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
AUTHORS      Pterygota: Neoptera: Endopterygota: Diptera: Brachycera:
REFERENCE    Muscomorpha: Ephydroidea: Drosophilidae: Drosophila.
JOURNAL      1 (bases 1 to 169526)
AUTHORS      Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
REFERENCE    Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
JOURNAL      Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
AUTHORS      Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
REFERENCE    Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
JOURNAL      Moshirefi, A.R., Moshirefi, M., Nixon, K., Pacleb, J.M., Park, S.,
AUTHORS      Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E.,
REFERENCE    Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
JOURNAL      Rubin, G.M.
TITLE        Sequencing of Drosophila melanogaster
REFERENCE    Unpublished
AUTHORS      2 (bases 1 to 169526)
JOURNAL      Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
AUTHORS      Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
REFERENCE    Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
JOURNAL      Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
AUTHORS      Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
REFERENCE    Moshirefi, A.R., Moshirefi, M., Nixon, K., Pacleb, J.M., Park, S.,
JOURNAL      Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E.,
AUTHORS      Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
REFERENCE    Rubin, G.M.
TITLE        Direct Submission
JOURNAL      Submitted (16-MAR-1999) Drosophila Genome Center, Lawrence Berkeley
AUTHORS      Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT      On Jun 1, 1999 this sequence version replaced gi:4803907.
              For further information about this sequence, including its location
              and relationship to other sequences, please visit our sequence
              archive Web site (http://www.fruitfly.org/sequence/) or send email
              to dbd@fruitfly.berkeley.edu. All contigs in this submission meet
              the following cutoffs: length >= 400 bases, phrap computed error
              rate <= 1/10.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 3 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              * 1 733: contig of 733 bp in length
              * 734 813: gap of unknown length
              * 814 41210: contig of 40397 bp in length
              * 41211 41290: gap of unknown length
              * 41291 169526: contig of 128236 bp in length.
FEATURES
SOURCE       Location/Qualifiers
1..169526
/organism="Drosophila melanogaster"
/strain="Y: cn bw sp"
/db_xref="taxon:7227"
/clone="BACR48023 (D498) RPCI-98 48.0.23"
/chromosome="3"
/clone_1lb="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6"
/map="99A7-99B5"
BASE COUNT  47509 a 37630 c 37526 g 46700 t 161 others
ORIGIN
Query Match      7.5% Score 19: DB 35: Length 169526:
Best Local Similarity 100.0%: Pred. No. 2.1:
Matches 19: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 130 TTGATGCTTTGGGCGCCT 148
| | | | | | | | | | | | | | |

```

Db 81329 TTGATGCTTTGGGCGCCT 81311

RESULT 6  
HS544A11 113143 bp DNA PRI 28-JAN-1999  
LOCUS Human DNA sequence from clone 544A11 on chromosome 22q12.1-12.3.  
DEFINITION Contains ESTs, STS and GSSs, complete sequence.  
ACCESSION AL023281  
NID g3449132  
VERSION AL023281.1 GI:3449132  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 113143)  
REFERENCE  
AUTHORS Spragon, L.  
TITLE Direct Submission  
JOURNAL Submitted (20-AUG-1998) E-mail enquiries: humquerry@sanger.ac.uk  
COMMENT  
Submitted (20-AUG-1998) E-mail enquiries: humquerry@sanger.ac.uk  
Clone requests: clonerequests@sanger.ac.uk  
On Aug 21, 1998 this sequence version replaced gi:3281981.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
IMPORTANT: This sequence is not the entire insert of clone 544A11.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone B42B1 (275890) is at 21055 in this  
sequence. The true left end of clone 732E4 (AL008722) is at 113044.  
The true right end of clone B42B1 is at 62012.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome 22, constructed by the Sanger Centre Chromosome 22  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr22  
544A11 is from the human BAC library described in U-J. Kim et al.  
(1996) Genomics 34, 213-218. VECTOR: pELOBAC11.  
FEATURES  
source  
1. 113143  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/clone="544A11"  
/map="q12.1-12.3"  
1. .4047  
/note="LI repeat: matches 1341. .5390 of consensus"  
repeat\_region  
3900. .4787  
/note="L1PA2 repeat: matches 3. .891 of consensus"  
repeat\_region  
5114. .5173  
/note="10 copies 6 mer gtaggt 80% conserved"  
repeat\_region  
5606. .5678  
/note="MIR2 repeat: matches 68. .146 of consensus"  
repeat\_region  
6710. .6783  
/note="Alu repeat: matches 289. .1 of consensus"  
repeat\_region  
/note="L1MC2 repeat: matches 976. .1055 of consensus"  
repeat\_region  
complement(8557. .8584)  
/note="MER5A repeat: matches 28. .1 of consensus"  
repeat\_region  
8614. .8732  
/note="MIR repeat: matches 49. .167 of consensus"  
repeat\_region  
8733. .9031  
/note="AluX repeat: matches 3. .302 of consensus"  
repeat\_region  
complement(9766. .10333)

repeat\_region  
11345. .11482  
/note="MIR repeat: matches 568. .1 of consensus"  
repeat\_region  
13487. .13705  
/note="MIR2 repeat: matches 1. .137 of consensus"  
repeat\_region  
complement(13755. .13830)  
/note="MER5A repeat: matches 101. .26 of consensus"  
repeat\_region  
14597. .14880  
/note="AluX repeat: matches 1. .300 of consensus"  
repeat\_region  
14881. .14928  
/note="8 copies 6 mer gaagaa 81% conserved"  
repeat\_region  
complement(15117. .15914)  
/note="L1 repeat: matches 914. .133 of consensus"  
repeat\_region  
complement(15929. .16229)  
/note="AluX repeat: matches 301. .2 of consensus"  
repeat\_region  
complement(16427. .16700)  
/note="Alu repeat: matches 302. .37 of consensus"  
repeat\_region  
16705. .17109  
/note="LI repeat: matches 4941. .5371 of consensus"  
repeat\_region  
17033. .17903  
/note="L1MB3 repeat: matches 53. .931 of consensus"  
repeat\_region  
17922. .18223  
/note="AluX repeat: matches 1. .302 of consensus"  
repeat\_region  
18379. .18679  
/note="AluX repeat: matches 1. .300 of consensus"  
misc\_feature  
19179. .19715  
/note="match: GSS B14146"  
repeat\_region  
20860. .20899  
/note="20 copies 2 mer tt 83% conserved"  
repeat\_region  
22323. .22619  
/note="AluX repeat: matches 1. .296 of consensus"  
repeat\_region  
complement(23253. .23411)  
/note="AluX repeat: matches 294. .130 of consensus"  
repeat\_region  
complement(23411. .23709)  
/note="Alu repeat: matches 298. .1 of consensus"  
repeat\_region  
complement(23714. .23837)  
/note="AluX repeat: matches 133. .9 of consensus"  
repeat\_region  
complement(23878. .24177)  
/note="AluX repeat: matches 299. .1 of consensus"  
variation  
24302. .24304  
/note="clone CB42B1; att in this entry; substitution"  
repeat\_region  
complement(24565. .24646)  
/note="MIR repeat: matches 144. .62 of consensus"  
variation  
24730. .24731  
/note="clone CB42B1; ta in this entry; deletion"  
repeat\_region  
25199. .25331  
/replace="taa"  
/note="LI repeat: matches 4892. .5022 of consensus"  
variation  
25415. .25417  
/note="clone CB42B1; ata in this entry; substitution"  
repeat\_region  
26174. .26318  
/note="MIR2 internal repeat: matches 290. .431 of  
consensus"  
repeat\_region  
complement(27275. .27542)  
/note="LI repeat: matches 1981. .1700 of consensus"  
repeat\_region  
28719. .28830  
/note="MER3 repeat: matches 1. .113 of consensus"  
repeat\_region  
29224. .29302  
/note="MIR2 internal repeat: matches 547. .625 of  
consensus"  
repeat\_region  
complement(29431. .29551)  
/note="MER20 repeat: matches 218. .98 of consensus"  
repeat\_region  
29557. .29912  
/note="L1MC3 repeat: matches 987. .1354 of consensus"  
repeat\_region  
29751. .29912  
/note="MER42c repeat: matches 2. .168 of consensus"  
repeat\_region  
30051. .31020  
/note="MER42b repeat: matches 166. .1204 of consensus"  
repeat\_region  
30531. .31246  
/note="MER42c repeat: matches 758. .1538 of consensus"  
repeat\_region  
32560. .32879

```

variation
    /note="AluJo repeat: matches 1. .301 of consensus"
    32705. .32717
    /note="clone CB42B1; aaaaaaaaaa in this entry;
    insertion"
    /replace="aa"
repeat_region
    complement(33362. .33507)
    /note="MIR2 repeat: matches 145. .1 of consensus"
    33749. .34016
    /note="match: EST AA313175"
repeat_region
    complement(34879. .34977)
    /note="MIR repeat: matches 203. .120 of consensus"
    35756. .35881
    /note="MIR repeat: matches 127. .256 of consensus"
    complement(36095. .36272)
    /note="MIR repeat: matches 210. .28 of consensus"
    38628. .38866
    /note="MIR repeat: matches 1. .262 of consensus"
    39980. .40309
    /note="MIR2 repeat: matches 13. .344 of consensus"
    4113. .41115
    /note="clone CB42B1; ata in this entry; substitution"
    /replace="aca"
repeat_region
    41142. .41177
    /note="MIR repeat: matches 84. .119 of consensus"
    42967. .43104
    /note="MIR2 repeat: matches 1. .141 of consensus"
    43327. .43627
    /note="AluSx repeat: matches 1. .302 of consensus"
    44907. .45200
    /note="AluSp repeat: matches 3. .298 of consensus"
    45200. .45201
    /note="clone CB42B1; ag in this entry; deletion"
    /replace="aag"
prim_transcript
    45385. .45626
    /note="match: EST T91723"
repeat_region
    45648. .45943
    /note="AluSg repeat: matches 5. .301 of consensus"
    46014. .46225
    /note="L1 repeat: matches 4565. .4778 of consensus"
    46262. .46561
    /note="AluJo repeat: matches 1. .300 of consensus"
    46573. .47151
    /note="L1M4 repeat: matches 208. .789 of consensus"
    47150. .47218
    /note="L1M3 repeat: matches 834. .898 of consensus"
    47242. .47285
    /note="22 copies 2 mer at 98% conserved"
    47244. .47285
    /note="7 copies 6 mer atata 100% conserved"
    47280. .47285
    /note="clone CB42B1; atatat in this entry; insertion"
    /replace="at"
variation
    47414. .47416
    /note="clone CB42B1; gtt in this entry; substitution"
    /replace="gct"
repeat_region
    48149. .48393
    /note="L1M2 repeat: matches 77. .334 of consensus"
    48407. .48436
    /note="15 copies 2 mer gt 100% conserved"

Query Match
Best Local Similarity 7.1%; Score 18; DB 9; Length 113143;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 CAGTGCATGATGCGACA 239
      |||
Db 75407 CAGTGCATGATGCGACA 75424

RESULT 7
LOCUS *AC004099 136222 bp DNA PRI 25-SEP-1998
DEFINITION Homo sapiens chromosome 17, clone HCIT42IK24, complete sequence.

```

```

ACCESSION AC004099
NID 93650064
VERSION AC004099.1
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutharia; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 136222)
Bairren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone HCIT42IK24
Unpublished
2 (bases 1 to 136222)
Bairren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckertly,R.,
Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J.,
Depayre,E., Devon,K., Dewar,K., Duretto,B., Ettemad,S.,
Ferreira,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S.,
Gerraldery,K., Gilmartin,T., Gray,D., Hagos,B., Harris,K.,
Horton,L., Howland,J.C., Hul,L., Jacotot,L., Linton,L.,
Mackenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrim,J.,
Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T.,
Paylin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D.,
Rollins,G., Rossello,R., Roy,A., Shyam,R., Sochoo,S., Stillwell,J.,
Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A.,
Wagner,A., Wheeler,J., Wu,Y., Ye,W.J., Zemtseva,I., Zhao,J. and
Zody,M.
Direct Submission
Submitted (01-FEB-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 136222)
Bairren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Boatin,C.,
Boutwell,C., Byrne,S., Cantu,C., Cerny,J., Collange,M., Collins,S.,
Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Duretto,B.,
Ettemad,S., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Gensheimer,S., Gerraldery,K., Gilmartin,T.,
Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hul,L.,
Jacotot,L., Kann,L., MacDonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nachman,A., Naylor,J., Naylor,J., O'Connor,T.,
Paylin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A.,
Shyam,R., Stange-Thomann,N., Stillwell,J., Stone,C., Strickland,C.,
Subramanian,A., Toruella-Miller,I., Vassiliev,H., Vo,A.,
Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (09-MAY-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 136222)
Bairren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Collange,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Gerraldery,K., Grant,G., Hagos,B., Heaford,A.,
Heinen,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lenoczky,J., MacDonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Naylor,J., Naylor,J., Naylor,J., O'Connor,T.,
O'Donnell,P., Paylin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Toruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (25-SEP-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 25, 1998 this sequence version replaced gi:126780.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html.
Location/Qualifiers
1. .136222
FEATURES
source

```



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HCT1421K24"
/clone_11b="Research Genetics/Cal Tech CTRB-HSP-C (plates
195-384)"
/map="17"
/chromosome="17"
257..525
/rpt_family="AluJb"
repeat_region
/rpt_family="AluJb"
complement(1015..1189)
/rpt_family="AluSg"
repeat_region
complement(1190..1487)
/rpt_family="AluSg"
complement(1488..1624)
/rpt_family="AluSg"
complement(1626..1667)
/rpt_family="(TAA)n"
complement(1668..1944)
/rpt_family="AluSx"
complement(1958..2074)
/rpt_family="L1MC4"
2101..2380
/rpt_family="AluJo"
complement(2882..3191)
/rpt_family="AluSg"
3295..3502
/rpt_family="AluSg"
complement(3503..3614)
/rpt_family="AluSc"
3619..3673
/rpt_family="Alu"
complement(3692..4000)
/rpt_family="AluSp"
complement(4002..4323)
/rpt_family="AluY"
complement(4479..4663)
/rpt_family="MIR"
complement(4796..5076)
/rpt_family="AluJo"
5097..5124
/rpt_family="AT-rich"
complement(5128..5283)
/rpt_family="FRAM"
complement(5338..5392)
/rpt_family="MIR"
5534..5833
/rpt_family="AluSg"
5847..5999
/rpt_family="FRAM"
6842..7136
/rpt_family="AluSg1"
complement(7291..7526)
/rpt_family="MTR1F"
complement(7555..7621)
/rpt_family="MTR1D"
complement(7741..7908)
/rpt_family="AluJ"
7909..8217
/rpt_family="AluY"
complement(8218..8233)
/rpt_family="AluJ"
complement(8266..8382)
/rpt_family="AluJ"
complement(8395..8502)
/rpt_family="MTR1D"
complement(8572..8724)
/rpt_family="AluJb"
complement(9729..9842)
/rpt_family="MER5B"
complement(10333..10478)
/rpt_family="MTR1F"

```

```

repeat_region complement(10627..10706)
/rpt_family="L1MC5"
repeat_region complement(10707..11009)
/rpt_family="AluSx"
repeat_region complement(11010..11093)
/rpt_family="L1MC5"
complement(11166..11299)
/rpt_family="L1ME2"
repeat_region complement(11300..11608)
/rpt_family="AluSx"
repeat_region complement(11609..12072)
/rpt_family="L1ME2"
complement(12070)
/notes="Probably 28 As. Three clones read 28 As, one clone
reads 27 As."
repeat_region complement(12093..12213)
/rpt_family="L1MC5"
12229..12273
/rpt_family="AT-rich"
repeat_region complement(12280..12396)
/rpt_family="AluJb"
complement(12414..12520)
/rpt_family="L1MD3"
12521..12814
/rpt_family="AluJo"
12816..13111
/rpt_family="AluSx"
complement(13112..13455)
/rpt_family="L1MD3"
13470..13596
/rpt_family="MER21B"
complement(13603..13748)
/rpt_family="L1MC5"
complement(13837..14111)
/rpt_family="L1MC/D"
14112..14284
/rpt_family="AluSg"
complement(14285..14409)
/rpt_family="L1MC4"
complement(14413..14450)
/rpt_family="(CAA)n"
complement(14451..14733)
/rpt_family="AluY"
complement(14736..15025)
/rpt_family="AluSg"
complement(15026..15160)
/rpt_family="AluJb"
complement(15503..15804)
/rpt_family="AluSg"
complement(16223..16536)
/rpt_family="AluY"
complement(16539..16832)
repeat_region

Query Match 7.1%; Score 18; DB 11; Length 136222;
Best Local Similarity 100.0%; Pred.No. 7.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 TGACAGGAGGTTTACAGA 93
Db 130510 TGACAGGAGGTTTACAGA 130527

RESULT 8
LOCUS RNU36785 1386 bp mRNA ROD 12-MAR-1996
DEFINITION Rattus norvegicus putative pheromone receptor VNI mRNA, complete
cds
ACCESSION U36785
NID 91039469
VERSION U36785.1 GI:1039469
KEYWORDS Norway rat strain-Sprague-Dawley.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus

```

REFERENCE  
AUTHORS  
TITLE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 1386)  
Dulac, C. and Axel, R.  
A novel family of genes encoding putative pheromone receptors in  
mammals

JOURNAL  
MEDLINE  
66028094  
2 (bases 1 to 1386)

REFERENCE  
AUTHORS  
TITLE  
Dulac, C. and Axel, R.  
Direct Submission

JOURNAL  
Submitted (20-SEP-1995) Catherine Dulac, Biochemistry and Molecular  
Biophysics, Howard Hughes Medical Institute Columbia University,  
701 West 168th Street, New York, NY 10032, USA

FEATURES  
source  
Location/Qualifiers

## CDS

1. 1386  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/tissue\_type="vomeronasal organ"  
/sex="female"  
270..1217  
/note="putative; seven transmembrane domain receptor."  
/codon\_start=1  
/product="pheromone receptor VN1"  
/protein\_id="AA92007.1"  
/db\_xref="pid:g1039470"  
/db\_xref="gi:1039470"  
/translation="MNKNSRLYTDNINTEFAEIGVANSLLLEIFKLCQ  
RSRLDPLIGLILMLMLAFATDTFISRMGMDDIICKSLIKYRTGRSLC  
TSLSLVALIILSPRSCLAKPKHSHSCALISLVLMFISHLVSTIATPN  
LTNDPIHYTOWCSILPMSTLMSSTLLATRDVPLISLMTLSITMYALLCRKRO  
KTHLOSTLSFPAKSPQRATRTSLIMASLFTVMSFDSIVCSSRTYMLNDPLISYIOL  
FMHIVATVSPFVITETKHIVNSKCVKVECLNIP"

BASE COUNT  
357 a 322 c 271 g 435 t 1 others

Query Match  
Best Local Similarity 100.0%; Score 18; DB 12; Length 1386;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 TCTGAGCTCAAGACTA 59  
|||||  
Db 1042 TCTGAGCTCAAGACTA 1059

RESULT 9  
LOCUS RRRASGRF 3804 bp mRNA ROD 03-DEC-1993  
DEFINITION R. rattus mRNA for guanine nucleotide releasing factor for Ras p21.  
ACCESSION X67241 S40901  
NID 957664  
VERSION X67241.1 GI:57664  
KEYWORDS guanine nucleotide releasing factor; releasing factor.  
SOURCE black rat.  
ORGANISM Rattus rattus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS  
TITLE  
Felig, L.A.

JOURNAL  
Submitted (07-JUL-1992) L.A. Felig, Tufts Univ. School of Medicine,  
136 Harrison Ave., Boston MA 02111, USA

REFERENCE  
AUTHORS  
TITLE  
Shou, C., Farnsworth, C.T., Neel, B.G. and Felig, L.A.

JOURNAL  
Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing  
factor for Ras p21  
Nature 358 (6384), 351-354 (1992)  
92350260

FEATURES  
MEDLINE  
source  
Location/Qualifiers  
1..3804  
/organism="Rattus rattus"

## CDS

/db\_xref="taxon:10117"  
/dev\_stage="adult"  
/tissue\_type="brain"  
/clone\_id="lambda gtl0"  
70..3804  
/codon\_start=1  
/product="p140 Ras-GRF"  
/protein\_id="CAA47666.1"  
/db\_xref="pid:g57665"  
/db\_xref="gi:57665"  
/db\_xref="SWISS-PROT:P28818"  
/translation="MQKIRLNDGHVAVSLGLAARDGTRKGYLSKRSSDNPKWTKWF  
ALLQMLLYEFSSSSRPSSGLYLLEGSICKPKSPKRTSGSKEDKHQVETVNFSD  
SOKSLIEDTDSKDCDEWVAIARASYIILTEAEHALQKLIHLQVETKTVAKOL  
ROOLEGDEVEIERLKAETANLIKONERIOSQLVAEPEDEDIKIKKVOSEFLGWL  
RKRMNIIDYIRSPHADMKNRQVMSVLEAEVVOCHIIIVNNPRLPMAAS  
KKPTTHDVSIFLNSFTTFELHQIFQGLKATLASPTVYVLDLIDILPMINTIO  
EVRNHQYSLQILAKCKONRDFDLKLOYEAKPDCSEETLETFILTPFQIPRYITL  
HELLAHPHEVERNSLDYAKSLLELSRVNHDSETEENRKNMLAERMTEGCEL  
LDTSQTEVRQGLILOVPMSEKGIKNGKLSLSLKGEGRCFLSKHLITCTRGS  
KLHLKNGVILSIDCTLLDDEPNDDGKGQEVHDLEKIVERKDSPPFVILVASS  
ROEKAAVTSDIICVDNIRCMGLMMAAFEEESKVTVMQIKSDASLCCDDVITFST  
MNSKCVQIRASVSRILRLTDLRPLSIDPLNFTLSYRFTTAVVYVLDKLSITIK  
PIPIAPARSIELLSSSRNKTLTGDAKPSRASKTSPPPLAIGTSSPYRRKLSL  
NIPITIGKALELASLGCPSDGYTNINSHPSFETKTLTDSKLVASSLTPPEIDM  
TLLESSEGRFRTSDILKEESDDQSDVDTEVFPPTPKSRNRTIOTEFPLFNNSGI  
MATCRDLMSNRSLATSAFAIATAGANESPAKEIYRRASLANTGYSPQRIKDE  
FVIRRAATNRVLANLRHVYKHSQDEFDDLLKVKICFLEEVHDDPLDPOEKAA  
NIMRLTDEEITEHNSMDLMLTEGKTEPPENHSAMETAEQITLIDHLYVFSTIY  
EEFGGGMKAKNERPTIYIMKTRHNSINLISLREYASASTERKAVANA  
DICIKAHYNAVLTITSSINSAIFRLKTKYKSKTKSLFDLQVLVSDGDFKRL  
RETRNCDDPVPVLPGLYLDLAFLEGTPTVTEGVLNFSKMMISHIILREIROFOO  
TYRIEPOPKVOTYLVDETFVLDESLEYEASLRLEPKLP"  
238..461  
/note="Db1 oncogene-like"  
misc\_feature  
920..1221  
/note="S.cerevisiae CDC-25-like"  
BASE COUNT  
1015 a 1077 c 932 g 780 t

Query Match  
Best Local Similarity 100.0%; Score 18; DB 12; Length 3804;  
Matches 18; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 55 AACTAGAGAGCTGTCCA 72  
|||||  
Db 1313 AACTAGAGAGCTGTCCA 1330

RESULT 10  
LOCUS AC006427/c DNA HTG 08-JUN-1999  
DEFINITION Homo sapiens chromosome 4, WORKING DRAFT SEQUENCE, 12 unordered  
pieces.  
ACCESSION AC006427  
NID 95019269  
VERSION AC006427.4 GI:5019269  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
Stone, N.E., Schmutz, J.J., Cox, D.R. and Myers, R.M.

JOURNAL  
Direct Submission  
Unpublished  
2 (bases 1 to 190000)

REFERENCE  
AUTHORS  
TITLE  
Stone, N.E., Schmutz, J.J., Cox, D.R. and Myers, R.M.  
Direct Submission  
Submitted (21-JAN-1999) Department of Genetics, Stanford Human  
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA  
On Jun 8, 1999 this sequence version replaced gi:4966377.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1122: contig of 1122 bp in length  
\* 1123 2793: gap of unknown length  
\* 2794 3669: contig of 1176 bp in length  
\* 3670 5639: gap of unknown length  
\* 5640 7494: contig of 1855 bp in length  
\* 7495 9165: gap of unknown length  
\* 9165 11251: contig of 2087 bp in length  
\* 11252 12921: gap of unknown length  
\* 12922 24188: contig of 11267 bp in length  
\* 24189 25859: gap of unknown length  
\* 25860 33032: contig of 7173 bp in length  
\* 33032 34702: gap of unknown length  
\* 34702 46042: contig of 11340 bp in length  
\* 46042 47712: gap of unknown length  
\* 47712 67667: contig of 19956 bp in length  
\* 67668 69338: gap of unknown length  
\* 69338 89177: contig of 19840 bp in length  
\* 89178 90849: gap of unknown length  
\* 90849 118409: contig of 27562 bp in length  
\* 118410 120079: gap of unknown length  
\* 120080 148176: contig of 28097 bp in length  
\* 148177 149846: gap of unknown length  
\* 149847 190000: contig of 40154 bp in length.  
Location/Qualifiers  
1. 190000  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"

BASE COUNT 52922 a 35299 c 34931 g 48469 t 18379 others  
ORIGIN

Query Match 7.1% Score 18; DB 35; Length 190000;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 TCATCAGGAATGAGCT 167  
|||||  
Db 178254 TCATCAGGAATGAGCT 178237

RESULT 11  
AC006034/c 56479 bp DNA PRI 14-MAY-1999  
DEFINITION Homo sapiens clone NH0144M13, complete sequence.  
AC006034  
NID 94827318  
VERSION AC006034.2 GI:4827318  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
TITLE 1 (bases 1 to 56479)  
JOURNAL Waterston, R.H.  
REFERENCE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 56479)  
TITLE Waterston, R.H.  
JOURNAL Direct Submission  
REFERENCE Submitted (22-NOV-1998) Genome Sequencing Center, Washington  
AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 3 (bases 1 to 56479)  
AUTHORS Waterston, R.H.

TITLE Direct Submission  
JOURNAL Submitted (14-MAY-1999) Genome Sequencing Center, Washington  
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

COMMENT On May 14, 1999 this sequence version replaced gi:1999982.

FEATURES  
Source Location/Qualifiers  
1. 56479  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NH0144M13"

BASE COUNT 18089 a 9723 c 10383 g 18284 t  
ORIGIN

Query Match 7.1% Score 18; DB 42; Length 56479;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 CTTTCATCAGGAATGGA 164  
|||||  
Db 15673 CTTTCATCAGGAATGGA 15656

RESULT 12  
BS0B0005/c 208430 bp DNA BCT 26-NOV-1997  
LOCUS Bacillus subtilis complete genome (section 5 of 21): from 802821 to  
DEFINITION 1011250.  
ACCESSION 299108 AL009126  
NID 9263055  
VERSION 299108.1 GI:2633055  
KEYWORDS  
SOURCE Bacillus subtilis.  
ORGANISM Bacillus subtilis

REFERENCE Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
AUTHORS Bacillus.  
1 (bases 1 to 208430)

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,  
Azevedo, V., Bortner, M.G., Bessières, P., Bolotin, A., Borchert, S.,  
Boriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,  
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,  
Carter, N.M., Choi, S.K., Codani, J.J., Comercon, I.F., Cummings, N.J.,  
Daniel, R.A., Denicola, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,  
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,  
Fougeron, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,  
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,  
Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,  
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hult, M.F.,  
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,  
Klaier, Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,  
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,  
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,  
Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P.,  
Mizuno, M., Mosetti, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,  
Ogawa, K., Ogilwara, A., Oudega, B., Park, S.H., Patro, V., Pohl, T.M.,  
Portetelle, D., Portwoll, S., Prescott, A.M., Priesack, E., Pujic, P.,  
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,  
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T.,  
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,  
Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B.,  
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,  
Takeuchi, M., Yamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,  
Tosato, V., Uchiyama, S., Vandendol, M., Vannier, F., Vassart, A.,  
Viari, A., Wambuti, R., Medler, E., Medler, H., Weitzinger, T.,  
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasunoto, K., Yata, K.,  
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and  
Danchin, A.  
The complete genome sequence of the gram-positive bacterium  
Bacillus subtilis  
JOURNAL Nature 390 (6657), 249-256 (1997)  
MEDLINE 98044033  
REFERENCE 2 (bases 1 to 208430)  
AUTHORS Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.

TITLE  
JOURNAL

Direct Submission  
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,  
Regulation de l'Expression Genetique, 28 rue du Docteur Pasteur, 75724  
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,  
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45  
68 89 48

FEATURES  
source

Location/Qualifiers  
1.208430

/organism="Bacillus subtilis"

/strain="168"

/db\_xref="taxon:1423"

/complement(5..1234)

/gene="yfnC"

/complement(5..1234)

/gene="yfnC"

/function="unknown"

/note="similar to fosmidmycin resistance protein"

/codon\_start=1

/transl\_table=11

/protein\_id="CAB12561.1"

/db\_xref="PID:e1182722"

/db\_xref="PID:g2633056"

/db\_xref="GI:2633056"

/db\_xref="SPTREMBL:O06481"

/translation="MAIAPLKEKTVOKRGTVVPIIIIGICHLNDISLAVPAM  
PLERSMSTFTGILGIIAFTLNMSSVWQVAVTDRKRPRIALPVGLTSMGLIG  
LAFASFTIIICVFICIGSAIFPESGSAVMAAGTRKRAQSIYVGGSGGAMA  
PIITALLVPLSQFGAVMPTLVAALAVMEKMTIAKMYASRLSLQKSGKOKKIA  
AITKSVSLIITLIFLIFARSWYTSAINETFYFAMDTYHSLIQOQSIYVFLFGA  
IGTFLGGPLADRFGRFVILGSLCSAPALVLPAGPVLAAGVALIGLIVLMSFVS  
TVVYQELVPGKIGTMSGLTVGLAFMGALGVALGALIDAAGLPTMTAIAFLPVLS  
ILAFLPSDOKLRHNS"

/complement(1345..2052)

/gene="yfnB"

/complement(1345..2052)

/gene="yfnB"

/function="unknown"

/note="similar to hypothetical proteins"

/codon\_start=1

/transl\_table=11

/protein\_id="CAB12562.1"

/db\_xref="PID:e1182723"

/db\_xref="PID:g2633057"

/db\_xref="GI:2633057"

/db\_xref="SPTREMBL:O06480"

/translation="MKRYRTLLFDVDDTLLDROAAEALALPLFEDONIPITNMKAO  
YKTIINGLMRAPEBCKMTREYVNTREPSALLKREYEDGALLKQKTRFLSEHOLI  
DGAEDLISNLQOQEDLYLVTVNGVSHTOYKRLKDSGLPEFFRDIFVSEDTGQPKMEY  
FNIVFERLPQFSAEHTLLIGDSLADIKGQLAGLDTCMNPNVPELIIPTYEIR  
KLEELYHLINIENTVSC"

/complement(2110..3529)

/gene="yfnA"

/complement(2110..3529)

/gene="yfnA"

/complement(2144..3529)

/gene="yfnA"

/function="unknown"

/note="similar to metabolite transporter"

/codon\_start=1

/transl\_table=11

/protein\_id="CAB12563.1"

/db\_xref="PID:e1182724"

/db\_xref="PID:g2633058"

/db\_xref="GI:2633058"

/db\_xref="SPTREMBL:O06479"

/translation="MSSLFRKKPIETISAGSKSLARTISAPDLTLIGVIGTIGI  
FVITGYAATGAGPALITISFLAGLACALAFCAEFSSSIPISGSVSYSYVLTGL  
LAFILGMDLMEVYALSAVATGSSYFOSLAFGNLHIPALTAGAGSAGAGAFNP  
AAVILITLITIVSRGVKESTFNNVYILMKIALITLITIVIGYVKKDMSPPMPFM  
KGVLSATVFAVFAFGDAVNASSEVKPNOKNPVGIISALAVCTVLIVAVSLVLTG  
MMPYAKLVGDPVSFAFGDAAVAGIISVCAITIGITVMTLLVIAQVAILTRMSD  
GLDGLFAKYVPSKTPFRNTMLTGLIYAAGIAGFINGTALHLVNMGTILAFVISA  
VIYLRKHPEIKASFRVFPVVPITISAGICLWMTSLPGVTWLSFVIMIAVGLTYE

## gene

## CDS

## LYSRHSLINK"

3779..5236

/gene="yfnB"

3779..5236

/gene="yfnB"

/function="unknown"

/note="similar to benzaldehyde dehydrogenase"

/codon\_start=1

/transl\_table=11

/protein\_id="CAB12564.1"

/db\_xref="PID:e1182725"

/db\_xref="PID:g2633059"

/db\_xref="GI:2633059"

/db\_xref="SPTREMBL:O06478"

/translation="MFOYELNKOFIGGKMOEGSSPVLENKNPYOKTFTTRKATA  
DDVEAYRAALAKKWDVAVNPEKRIIEKAVYIENNEALIIYILMEELGSTRKA  
AFETGIVKNIKAKATPIIREKTIESTIDGKRNRLRYPAAGVGVISFNPFFLS  
KSVAPALGANGVVLKPHETPICGGLTIAKLEENAGIPAGLVVVDIAIGDSF  
VEHPVPRIISFTGSTKVGSIQOLAMKFRKPLLELGNSAFVLEADIEVAVNAV  
FSRPTHOGICMSANRVLVHSSIVDKFLELQAKVESLKGDPMDPTIIGPLVNSHO  
TDGLMKTVEQAIIEGAVPVLKGFNGTIVETILKDYKPMNSIAKEELFEPVSPKFE  
DSEDEAVDIANETPFGSAVHTSNLGRVAFAPAKRIETGWHIYNDTTINDEPVAATG  
EKOSGLRNGCENSELEPTILNKISVQHEKRSPT"

5250..6136

/gene="yfnB"

5250..6110

/gene="yfnB"

/function="unknown"

/note="similar to methyl-accepting chemotaxis protein"

/codon\_start=1

/transl\_table=11

/protein\_id="CAB12565.1"

/db\_xref="PID:e1182726"

/db\_xref="PID:g2633060"

/db\_xref="GI:2633060"

/db\_xref="SPTREMBL:O06477"

/translation="METITTEKETADILDAPAKVAPYLSLVODITIGIDTEKL  
VNIPAKTFLSNRAGDPLOGEDLITDAIRSNOKTMRVPELFGPLIARALPLHEN  
GRVIGVGLTSLNESSKLDHVAESLSAVEQTAASIDISEINSGFSTOMGISOA  
KRVESAGEIADISVTVKGISDSNGLGNAALAEARAGESGKGFVADIEIKLATH  
SKENVOIQDITKRIHSLKGLSESIENQHTDQAAVQISATMTQESISQAOLA  
KMAEKLEEE"

6113..6136

/gene="yfnB"

6245..8159

/gene="yfnB"

6245..8134

/gene="yfnB"

/function="unknown"

/note="similar to ABC transporter (ATP-binding protein)"

/codon\_start=1

/transl\_table=11

/protein\_id="CAB12566.1"

/db\_xref="PID:e1182727"

/db\_xref="PID:g2633061"

/db\_xref="GI:2633061"

/db\_xref="SPTREMBL:O06476"

/translation="MSILKAENDIKYGGKTLFDHISFHEENRIGLIGPNTKST  
LTKVANGLESIEGELTKSGSVQVEFLHQDPLPAQVLYLEHISGESVMTKLERE  
KALYELGKDPENBOROKHLLAAQAKDANNANDAMTIAATVSKGVNDVTPVNELS  
GGOKRVAIAKNILOPADLLIDEPNHLNDNETIMLEGYLSQYGAVLVTHDRVFL  
NRYNRIYELERGLSYTYKGNVEFLERAEAEQAEQETROULRELALMRGA  
KARSTOKARIDRVEIKQOTPOSSGSIDPRIGSHRUKOYIEENWIAIDGRMLV  
DRNELVYGERIIGIPNGIGKTTLLNLAGRHHPDDGDTITIGTVAGYTTQDSE  
MNGELKVIDYIKETALEVNTADMDITAEOMLERLFRSMQOYIRKISGGERRLY  
LQVLMQENNVFLDEPTNDLDETLSYLEDYIDQFPGVIVTSKDRVLDVDRVRLI  
VEEGQVIRFOGYSYDVEESKAKAKAPPAEKRTAEAEKRRKRLSKYDQLEMD  
GIEDKIAQOEKHEHLENDIAAGSDPGKIQELMAEQAKTAELLEAAADRTELSMT  
ELES"

8138..8159

/gene="yfnB"

8257..8740

/gene="yfnC"

terminator

gene

```
CDS      8257..8703
          /gene="yfmQ"
          /function="unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB12567.1"

Query Match      6.7%; Score 17; DB 1; Length 208430;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      167 TGCCTCTCAGCGCTCCA 183
Db      99303 TGCTCTCAGCGCTCCA 99287

RESULT 13
LOCUS   D78508      9444 bp      DNA      BCT      05-FEB-1999
DEFINITION Bacillus subtilis DNA for yf10, yf1P, yf1N, yf1M, yf1L, yf1K, yf1J,
            yf1I, yf1H, complete cds.
ACCESSION D78508
VERSION   91817531
KEYWORDS  yf1H; yf1J; yf1K; yf1L; yf1M; yf1N; yf1O; yf1P.
SOURCE    Bacillus subtilis (Strain:AC327) DNA.
ORGANISM  Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
            Bacillus.
REFERENCE 1 (bases 1 to 9444)
AUTHORS   Sekiguchi,J.
TITLE     Direct Submission
JOURNAL   Submitted (27-NOV-1995) to the DDBJ/EMBL/GenBank databases. Junichi
AUTHORS   Sekiguchi, J., Textile Science and Technology, Shinsu University,
TITLE     Department of Applied Biology, 3-15-1 Tokida, Ueda-shi, Nagano 386,
REFERENCE 2 (bases 1 to 9444)
AUTHORS   Sekiguchi,J.
TITLE     Nucleotide sequence of Bacillus subtilis chromosome around 77
REFERENCE 3 (sites)
AUTHORS   Yamamoto,H., Uchiyama,S. and Sekiguchi,J.
TITLE     The Bacillus subtilis chromosome region near 78 degrees contains
JOURNAL   the genes encoding a new two-component system, three ABC
AUTHORS   transporters and a lipase
REFERENCE 97128783
JOURNAL   Gene 181 (1-2), 147-151 (1996)
FEATURES
     source
     location/Qualifiers
         1..9444
            /organism="Bacillus subtilis"
            /strain="AC327"
            /db_xref="taxon:1423"
         102..107
            /codon_start=1
            /transl_table=11
            /product="yf1H"
            /protein_id="BA11398.1"
            /db_xref="PID:d1012063"
            /db_xref="GI:1817532"
            /db_xref="GI:1817532"
         117..1058
            /translation="MKLSYVDSLGLHPEDMLNPAKLGIDLTLEMTGWSAPAHNL
            IDELQSEKREKFSQALEKRNATICALNSGNPLDPELGKSHRIDTKMELAGL
            GYKRYIMSGPAGGDDKVPWITVYVSMPVLDILNOMEDVAIPYKRLVKAE
            AGCVKIALNENSSOLYNNPFLRLRNAGVPMGINDPSHLMGADPTIAAEELS
            SAIHVHGKDVIERILSKVNGLETKETIDPANRANMTVAAGCGDQDMKEFTSV
            KMGVEGVSELMEDLTMSPEAGIRTSVALKQTISO"
         1055..1063
            /transl_table=11
         1068..2249
            /codon_start=1
            /transl_table=11
```

```
terminator
RBS      2263..2290
CDS      2360..2368
          /product="yf1I"
          /protein_id="BA11399.1"
          /db_xref="PID:d1012064"
          /db_xref="PID:91817533"
          /db_xref="GI:1817533"
          /translation="MNLGERLISKPLRMAYVGGRLSVOGYKHRIGALRDNTAROLT
          AGAPIDAEKRGDEGVNNGVDAERCEYPTQWTFAEARROGIEVVSATPMTXKEI
          CNALEAGVHYICEKPLFTSNAGEILAKERKQIVGTTIGSGNOLMARMATE
          LQMDIRIVVDLQYTHGFCATDEGERISAQKWRVDPALAGPSFVGLDLSHTTYMSO
          LIMPDKIRIELCDROSFVSGRAPLEDNAYLMHYNCAVGMTSTSIAGCMGCHRI
          RIVGSKASIEEMWDSKPNELTYEVOGEPQOTLVRGMPYLDVYCNADERLGALHSELSLSE
          AMANILKPAIAIDAKNREDEILNNLVYPIDAGIEGIRMIENCVRSANQSVWVEF
          K"
          2263..2290
          2360..2368
          2373..3575
          /codon_start=1
          /transl_table=11
          /product="yf1J"
          /protein_id="BA11400.1"
          /db_xref="PID:d1012065"
          /db_xref="PID:91817534"
          /db_xref="GI:1817534"
          /translation="MKALFTTMTFTLWVSCMLYSIVKEDNMFGYVFIAAGAAVAA
          NYLAKRENATIFCLDIAIGFSRGFTIPGGLFTIMCPVAVAFELGFPKRTAMV
          LCLSLILFLVLRITAFMGNEFVIDHLSMTFVFCVGVGLKRLDADQDTAKQF
          QELTESHLSAHOELHLYAKOVDELTAIERNMAEIHDTVHKMTALVQQL
          REMQKRDQKADETVGCVETLREALDLYRVSRLQTEPNPSLIESKOLTEDCKN
          AGVTEFAVSGDPAITIPSLPHTLIRTVQEAITNKRGGAAACSIQACTDSISLV
          IRDQKGNPEALGFGGLNMKRAAHEGMIRFESERQGFVYVNEFSLANKMSEFGP
          VQKRESLS"
          3557..3563
          3572..4234
          /codon_start=1
          /transl_table=11
          /product="yf1K"
          /protein_id="BA11401.1"
          /db_xref="PID:d1012066"
          /db_xref="PID:91817535"
          /db_xref="GI:1817535"
          /translation="MKIITIDDDIVREGLASLLQREELDVITARNQGEAFERAK
          ELEPDIYMDIRMPVSNVGEKLTSSLPYKVMLTPTPDSALIAALEGASGL
          LKDMADITVAKMTHVHSGAVLPPELIRQMLNENKRRKQKIGTIEIKRPNLLDTE
          REIEVLALGIGLNKETAERKLYITEGIVKNHVSNTISKLAVRDTQAIIVSRIGVS
          VF"
          4265..4287
          4362..4368
          4378..5313
          /codon_start=1
          /transl_table=11
          /product="yf1L"
          /protein_id="BA11402.1"
          /db_xref="PID:d1012067"
          /db_xref="PID:91817536"
          /db_xref="GI:1817536"
          /translation="WLAENIKKAYGKRTIVKGISFSLKKESEFGLDPNGAKSTTI
          SMISGLVPHDSNITVGVYIGKEPAKKOKGIYPOPIALYPTLTAEHNLFWKMY
          GLTHDEAKRAAEVLEYGLTERAKDKLETFSGGKRIINIGALMLHMPLEIDPEPT
          VGIDPOSRNHILETVKOLNETGMTVITYSHYMEVEYFCDRGIIIDGEMALIGKTD
          LCSRIGDPTIQLTVSGINPAFIVAIRSLAVNDVTVADELKIDISAAHHEKVTSL
          LAERAHINILSLQVQEPNLERPLNLGTGLRD"
          5313..5318
          5326..6516
          /codon_start=1
          /transl_table=11
          /product="yf1M"
          /protein_id="BA11403.1"
          /db_xref="PID:d1012068"
          /db_xref="PID:91817537"
          /db_xref="GI:1817537"
          /translation="MKRSIWIAMKDVKIRITDRKGFNMLIMPLLTICILGALGSV
          DGSRIIDIKVGYIOSDOSDTANETKVLAKMKSIVTKYVSGDKMKKLIDKKRIDV
          GIVPNHWEAKRTSAVVAAPDQTLKSSIETAASSFIEQTKAYKEAAGSGMDYISKI
```

```

EAVKQCKLDPQAFKELATLEKETGDKLTLLIAEKSVSKAVTSFOYTSNAMLCEMLE
HITVGAKESELOKEDTETLARMMLTTPAKQSVILFGKWLGYLFALIOFFLPLVITNF
GYDWGNLILVSLGISTYAAVSGISMLASCSIDMKRTDADIGGGLIOLAVLGGSMF
PLYOFEDVLQSVSKAVPNRMALDGLSLMEGGWMDLQKPVLLFPAIGCSLVITGR
LHRR"
RBS      .6522
CDS      6530..7687
          /codon_start=1
          /transl_table=11
          /product="yfiN"
          /protein_id="BAA11404.1"
          /db_xref="PID:d1012069"
          /db_xref="PID:g1817538"
          /db_xref="GI:1817538"
          /translation="MKKILAIQGISLISLFKKPKQNYLIMFAPALLITFEVGSMLSGND
          DKVRLAIVDDDTILSOHYIRQLKADMDYVENMESKASEKLQKIKIAGIIVSRS
          FQTOLEKGNHPELIFRHGPELSEAPMVQVAYSALATINIOYTAKTASQTAGEMWKA
          AKYVEAKKHEDIVPAVROTISDKKEGAEADPTASRAAGSILFVMTMGAGACTIL
          EARKGVMSRLITASVRAEIGAGVLSFTFYGMIOFELISTHMLTGIMGNPAAY
          IYVLSPLITVIGIGLMTANVRIPEQDLQAGNLFIATCMVSGMWYPIEDPEFMOS
          IAEFLPQKMWASGLTEIILANGARVTDIIGIGIILAFAPAITFAAGLKALRA"
          7686..7719
          complement(7760..8308)
          /codon_start=1
          /transl_table=11
          /product="yfiO"
          /protein_id="BAA11405.1"
          /db_xref="PID:d1012070"
          /db_xref="PID:g1817539"
          /db_xref="GI:1817539"
          /translation="MRVLKVAIILGLRKELSGVDITSYFKEELGOFMSAKHSQIYEE
          LKLIKDEGIFERTIOGTLEKKWYITLDSGKQELHMLRHRPIETVYKDEMLKA
          VEIGLSQESADSEFKDQLQKQAKLSDLQGSYERKLWASAEPMSSSPDEHYLVTK
          ALERKNTVSWESTLAMDND"
          complement(8314..8318)
          8569..8573
          RBS      8581..9213
          CDS      /codon_start=1
          /transl_table=11
          /product="yfiP"
          /protein_id="BAA11406.1"
          /db_xref="PID:d1012071"
          /db_xref="PID:g1817540"
          /db_xref="GI:1817540"
          /translation="MKKVIAFFICLSLISVLAAPSGAKESVHPVYLVGICSA
          SYNEFAIKNTIYISGQMSNKIAXIDFDKRGNNINNGPOLASTYDRVLAKETGAKVDI
          VAHSGANTLITIKYIGGGKIKQNVVTLGANGVSTALPGTDPQKILITYSISL
          NDOIVINSLSLQAGARNIOLYIGIHGILNSOVNGYIKEGNLNGGLINTN"
          9224..9258
          terminator
          BASE COUNT 2654 a 1996 c 2376 g 2418 t
          ORIGIN

```

Query Match 6.7%; Score 17; DB 1; Length 9444;  
 Best Local Similarity 100.0%; Pired. No. 32;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 TGTCCTCAGCGCTCA 183  
 ||||||||||||||||  
 Db 1179 TGTCCTCAGCGCTCA 1163

RESULT 14  
 SYCSLRB 146271 bp DNA BCT 13-FEB-1999  
 LOCUS Synechocystis sp. PCC6803 complete genome, 19/27, 2392729-2538999.  
 ACCESSION D64000 AB001333  
 NID 91001484  
 VERSION D64000.1 GI:1001484  
 RACD: 16S rRNA: 23S rRNA: 3-hydroxyisobutyrate dehydrogenase: 30S  
 ribosomal protein S21: 50S ribosomal protein L20: 50S ribosomal  
 protein L35: 5S rRNA: ABC transporter: AbcL-like: ATP-dependent Clp  
 protease regulatory subunit: DNA recombinase: GMP synthetase: Mg2+

transport ATPase: RadC protein; UDP-3-o-[3-hydroxymyristoyl]  
 glucosamine n-acyltransferase; UDP-N-acetylglucosamine  
 1-carboxyvinyltransferase; aminophosphoribosylglutamine  
 aspartate aminotransferase; carboxyl-terminal processing protease;  
 cell division protein FtsH; cobinamide kinase; cytochrome b6;  
 cytochrome b6-f complex subunit 4; cytosine-specific  
 methyltransferase; ferredoxin; fructose-1,6-bisphosphate aldolase;  
 fumarate: glucose inhibited division protein A; glucose transport  
 protein; glucose-1-phosphate thymidylate transferase;  
 glutamate-1-semialdehyde 2,1- aminomutase; glycyl-tRNA synthetase  
 beta chain; high-affinity branched-chain amino acid transport  
 ATP-binding protein; histidyl-tRNA synthetase; lipid A disaccharide  
 synthase; mannosyl transferase; membrane-bound lytic  
 transglycosylase A; methylenetetrahydrofolate dehydrogenase;  
 oligopeptide transport system permease protein; penicillin-binding  
 protein 1A; photosystem II PsbT protein; plastocyanin; protease IV;  
 protein-export membrane protein SecD; protein-export membrane  
 protein SecF; protochlorophyllide reductase ChlB subunit;  
 ribonuclease D; ribonuclease III; ribulose biphosphate carboxylase  
 large subunit; ribulose biphosphate carboxylase small subunit;  
 sensory transduction histidine kinase; tRNA Ala; tRNA Gly;  
 tRNA-Ile; tRNA-Ser; thiol-specific antioxidant protein; thioresdoxin  
 M; transcriptional regulatory protein HlyF; transposase;  
 triosephosphate isomerase; unknown function; Ycf33; Ycf38 gene.  
 Synechocystis sp. (strain:PCC6803) DNA.  
 Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 1 (bases 1 to 146271)  
 Tabata,S.  
 Direct Submission  
 Submitted (30-NOV-1995) to the DDBJ/EMBL/GenBank databases. Satoshi  
 Tabata, Kazusa DNA Research Institute, Gene Structure 2; 1537-3  
 Yanauchino, Kikazaru, Chiba 292, Japan (E-mail:tabata@kazusa.or.jp,  
 Tel:0438-52-3933, Fax:0438-52-3934)  
 2 (bases 1 to 146271)  
 Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N.,  
 Sugita,M. and Tabata,S.  
 Sequence analysis of the genome of the unicellular cyanobacterium  
 Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 region from map positions 64% to 92% of the genome  
 DNA Res. 2 (4), 153-166 (1995)  
 3 (bases 1 to 146271)  
 Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y.,  
 Miyajima,N., Hirosewa,M., Sugita,M., Sasamoto,S., Kimura,T.,  
 Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naro,K.,  
 Okumura,S., Shimp,S., Takeuchi,C., Wada,T., Watanabe,A.,  
 Yamada,M., Yasuda,M. and Tabata,S.  
 Sequence analysis of the genome of the unicellular cyanobacterium  
 Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 entire genome and assignment of potential protein-coding regions  
 DNA Res. 3 (3), 109-136 (1996)  
 Potential protein coding regions were assigned on the basis of  
 similarity search of the ORFs and Genemark analysis.  
 FEATURES  
 source  
 JOURNAL  
 MEDLINE  
 COMMENT  
 CDS  
 1..675  
 /db\_xref="taxon:1143"  
 /note="ORF\_ID:slr0388"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAA10110.1"  
 /db\_xref="PID:d1010761"  
 /db\_xref="PID:g1001485"  
 /db\_xref="GI:1001485"  
 /translation="MWSIKSLIRITQPCPLQCRFSQDLCRDCTGQILSHRPFKFOQ  
 WQGNFPRVQGVQGLRRLATVTKFQGDIDIGLAEQALAEQALQOPAKTKLRPQV  
 VPIPLNRKQVQGFNOAERLAFGFCRLTGVLHPLOAKVFKDQALFEGSPGRRE  
 LOSALATGPGENTRRPMLIIDITTTTLEARNIMEKAGVILGVALAAPSFSOP

```

CDS
    METR00"
    /note="ORF_ID:s1r0769"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BA10111.1"
    /db_xref="PID:d1010762"
    /db_xref="PID:g1001486"
    /db_xref="GI:1001486"
    /translation="MOKVSPRLARVALMTGSLTATSLVTPPAOASSLSNGSAIP
    PKPIESTATVPAPATSPSPNTSQALATHVLYGEEKVYAQMDKRLASYPVAG
    KKGWETPGCNFOVIQWENPVENPNKRVASLDGPICIRNIGTWSGKNTIGRHG
    TKRKHHEFLGTAASHGCVARMRQDVALLNMVQNGTPRVQVK"
    1504. 1824
CDS
    /note="ORF_ID:s1r0770"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BA10112.1"
    /db_xref="PID:d1010763"
    /db_xref="PID:g1001487"
    /db_xref="GI:1001487"
    /translation="MATLILNHLEITPGCGGKPRAGHRITVODIVITHEEMGLSPD
    EIVHHPSTTLADVTAALSYHDHLEIRNGIIDEITPEMRSSITPSLMEAKLRVH
    GQD"
    1808. 2143
CDS
    /note="ORF_ID:s1r0771"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BA10113.1"
    /db_xref="PID:d1010764"
    /db_xref="PID:g1001488"
    /db_xref="GI:1001488"
    /translation="MANRIKFLHDENISIALGHRSLDVTTPETGLGVSQDO
    LOPASSEGRVITQDIDFLKAGSSDHRGIVACSRKNSIDQVIGLVAMEILEPE
    ECKNHIEFL"
    2375. 3901
    /gene="chlb"
    2375. 3901
    /gene="chlb"
    /note="ORF_ID:s1r0772"
    /codon_start=1
    /transl_table=11
    /product="protochlorophyllide reductase Chlb subunit"
    /protein_id="BA10114.1"
    /db_xref="PID:d1010765"
    /db_xref="PID:g1001489"
    /db_xref="GI:1001489"
    /translation="MKLAYMYAGPAHIGTLRIASSFKNVHAIMNAPLDDYFNVARS
    MLERONFPTVTSVVDROVLSRSGQEKVDNIVRKGSETPDLIVLPTCSLSIQE
    DIANFVDFRAOMDACHGVLLADYNHYRNELQAGDRIKQIYERIKKARQNLATKE
    TAKPSVNLGTFITIGHNOHDTLKRIMADGIEVNLILPEKATVDLAKIPQAMFN
    LCPYREIGLMTAFYLDSEFAQPYVDITPAGVETACIKRMOQVNLQSGFDVNEFI
    QOQTRHVSQAMFSSRIDCONLTKRAVFGDTTHAAATKILAREMGJHVVLAGTIC
    KYDAWFERAEVSEYCEVLISDNGAIAVALRIEPAALFTQMEHRAVKRLDIPGV
    IADPIHIOFPLGYKPFLEGTINOJADIVYNSFTGMDHLEITGSDTEVITKG
    ISSDSDLGWHSTAOAELNKVPGFVRGKVRNTEKFAREGLSEITLEVYAAKEAVGA
    "
    4033. 4713
CDS
    /note="ORF_ID:s1r0773"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BA10115.1"
    /db_xref="PID:d1010766"
    /db_xref="PID:g1001490"
    /db_xref="GI:1001490"
    /translation="MYILIGMGSMGSLAKNLKMGHTVAAVDTPPLACQYAREKIG
    VAFESGANTSVLLEAGVRKAGAVAAALQEDALNALVTLSEHYGVPIIVMSPGD
    FAPYRLAGATTHIISTELARVINAITYPOVDAMHDEGOIEVLAKPISPNCTIV
    GSVAAIADQPRFPSCITLITIGQANAHEDLKIPNGNTLLEDSTILAATKPELVRLI

```

```

CDS
    DEMGLQGS"
    /complement(4750..6390)
    /note="ORF_ID:s110772"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BA10116.1"
    /db_xref="PID:d1010767"
    /db_xref="PID:g1001491"
    /db_xref="GI:1001491"
    /translation="MGLICGLAASGLVTPISSEATQISVNEFPFVNVNIOGASNLG
    NGVNYVTSAELRDVQPSDMAALQASIVPERGCVGPDPRYRGDAGCTLEAPRLR
    YEFAGLANCLNTIQLQDENVSAQGDIDLKKAQOPALKALAYRVDLERTIRA
    FLEDDQFSTYTLIGQTTISLTKVFDVRGGSNEFQTFAYRFRNLETSFGKDLF
    RRLQFSNPLNGVEQGTQNTSFNDNSNNNVEVSHMYRPLNDLTLRLGTGVG
    YVDLDTLPLPTIADALGIPRFEGEYDPVYRGGGAGFNQIIPVQLSVGLATN
    PNDPDRGNGLFNGGYHTLAQLAYOTADGGIGFTYSRSYEPAGNTDLMGTSILAOP
    RGEVATAGNFTYITQGYRITPHNOIHMGGVVEAQQGGISNLNCGVGTPLRDS
    KNSAIWGLVGISFPDVGEGCNLPGIALGIPPTTASNLPGAVGQTPYHLEATYRI
    QLNDNISITPGVWVLANPENSNNATQYVGHRTSFLF"
    /complement(6482..7888)
    /gene="gicp"
    /complement(6482..7888)
    /gene="gicp"
    /note="ORF_ID:s110771"
    /codon_start=1
    /transl_table=11
    /product="glucose transport protein"
    /protein_id="BA10117.1"
    /db_xref="PID:d1010768"
    /db_xref="PID:g1001492"
    /db_xref="GI:1001492"
    /translation="MNPSSSPQSTANVKEVLLISGVAALGGFLGFDPAVINGAVAA
    LQKHFOETSLTGLSGLSALGSAAGAGPIADRHGIRKIMIAAVFTLSISGG
    LPTITMDFEFMRVLGIGVGAASVIAPIAVSPAHRLGRSLQOLAIVSGIFIAL
    LSNFRIAYMAGGSANPNLFGAAARMWETLIPALLYGCAFLIPESPRLVAVGO
    GERKAIIMKVGQVPSRIPRIOTVTLSDHPRSDLSLRSGGLPLTYWIGMSAL
    OQFVGINVITFYSVILMSVGTERTKSLIITITFTFNLITLVAAIPYDKRGKPL
    LMGSTGNTTIGLISLVFGATVVGQPLTGAAQIIVLAVNLKLVAFSFGSWGIIV
    VLLGEMFNKIRAAALSVAGVQWIANFIIISTFPPLDVLGPAVGLAYTASAIISI
    FPIPFVKEKTKGLEQM"
    8186. 9604
    /gene="secd"
    8186. 9604
    /gene="secd"
    /gene="secd"

Query Match      6.7%: Score 17; DB 1; Length 146271;
Best Local Similarity 100.0%: Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 GGGATTACCCCTCACA 216
DB 19590 GGGATTACCCCTCACA 19574

RESULT 15
BOVNEURXIA 6349 bp mRNA MAM 09-AUG-1995
LOCUS Bos taurus neuexin I-alpha, complete cds.
DEFINITION L14855
ACCESSION NID 9388560
VERSION L14855.1 GI:388560
KEYWORDS neuexin I-alpha.
SOURCE Bos taurus brain.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 6349)
AUTHORS Ullrich,B., Ushkaryov,Y.A. and Sudhof,T.C.
TITLE Cartography of neuexins: more than 1000 isoforms generated by
alternative splicing and expressed in distinct subsets of neurons
JOURNAL Neuron 14 (3), 497-507 (1995)

```

```

MEDLINE 95209856
FEATURES
source Location/Qualifiers
1..6349
/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue_type="brain"
1293..1382
/evidence=experimental
sig_peptide
1293..5885
/note="amino acid information: bases 1383 .. 5657
extracellular domain; amino acid information: bases 5355
bases 5658 .. 5717 transmembrane domain; amino acid
information: bases 5718 .. 5882 cytoplasmic domain;
precursor"
CDS
/codon_start=1
/product="neurexin I-alpha"
/protein_id="AA74123.1"
/db_xref="PIR:9388561"
/db_xref="GI:388561"
/translation="MGATLQGGCFLCLSLILGCAEELSGLEPPGAGOWTRFP
KMAACSEMSFOLKTRARGLIYFDEGFCDFELILTRGRQLQSFISCAEPAT
LITDTPVNDGAMHNRIRPFNTLFDIOYAKVEXSRKMDTPSGLVGLP
ELRAALKLTLASVREPERPKMIDVAVNSIALPVDGKIDDEPNSGGSPCE
AGEBEGCVCLNGCYCVYDQAVDCSRTGRCGCCQEDNNVGLAHLMGDOGKS
KEDNNVEGLAHLMGDOGKEETATFKSEYFCYDLSNPQSSDDETLSTKLQRLN
GLMHTGASADYVNLALNGAVSLVINGSGAFALVSPVNGKENDNMHVDKTYRLN
RQTSGIGHANVNLKHCSTYISVDGILTTGYOEDPKYILGSDPEYVGSSTADLP
SPVSNFMGCKLEVVYKNNVDYELSLAKOGDPMKIHGVAFKCEMVALDPTFE
TPESFISLPKNNAKTGISFDPERTENGILFESHGKPRHOKAKHPOMKVPEAF
EMDGHATLLDMGSGTIRKAKOKKVDGMVHYHDPDRSGSISYNTARTPTAT
GESQILDDELIVGLGLENKAGLVPIEVTALNTIGVIGRLFDGSKDRQM
AEVOSTAGVAKPSCSEETAKPCLSNPKCKNGMCRDWNVYVCDSCOTGLRSCREAT
VLSYDSGEMFKIQLEPVVMTAEVDYSLFRQRAIGIMATISDSADTLRLDELADR
VKLYVNLDCIINCNSKPEITLFAVYNNMNTVAVRGSILKLVDDQAMTG
OMAGHRTLEPHNIETGILTERRYLSVSPNIGHQSLTEGMAVIDLCKNGDIDYC
ELNARFGRNITADPVTFRKSSVALATLAVYFMHLPQFKTIDGLIYNSGCG
MDNIVELVKGYLHYFDPDNGANLIKSSSKPLDNDNMHNVISRDPSNHTYKIDT
KITQITGAKRNLKSDLYIGVAKETKSLPKVAKKEGQGLAVDINGRPLD
ISDALFCNGQIERGCEGSTCQEDSCNAGCQOWDQICDSMTSFGPLNDPG
TTLVFSKGGQITTKPNDPSTGRADRLAIGFSVQKAVLVRVDSGSGDYELH
IHQKIGVAFNVGTDIALIESNAHNGKHVAFTEGNAITLOVNSWVIEYPA
GNNONERLAIARORIPYRLGRVDEWILDKGROLITEGSOATIIIGKEGQOPRO
SGIYNGKAVIAMAENDANAIVGNVAVGEVBSMTTESTATRMQSEMTSMETT
TTLATSTARKRPPKEPVSOITDILVASACPSDDIDPCEPSSGGLANPTRAG
REPYPSGALEVIRESSTGTGMYGIVAAALCILLLYMYRYRNDEGSHYVDESRYN
ISNSAOSNGAVVKEKOPSSAKNKNKKEKEYV"
1383..5882
/function="presynaptic cell surface protein"
/evidence=experimental
repeat_region
1452..2699
/product="neurexin I-alpha"
/note="major repeat A; putative"
/rpt_family="neurexin repeat"
1452..1865
/note="left arm aa of the major repeat A; putative"
/rpt_family="neurexin repeat"
1941..2057
/note="EGF repeat 1; epidermal growth factor; putative"
/rpt_family="EGF"
2065..2172
/note="spliceable sequence, splice site 1"
/evidence=experimental
2262..2699
/note="right arm Ab of the major repeat A; putative"
/rpt_family="neurexin repeat"
2475..2519
/note="spliceable sequence, splice site 2"
/evidence=experimental
2871..4019
/note="major repeat B; putative"
/rpt_family="neurexin repeat"
2871..3314

```

```

/note="left arm Ba of the major repeat B; putative"
/rpt_family="neurexin repeat"
3375..3497
/note="EGF repeat 2; epidermal growth factor; putative"
/rpt_family="EGF"
3513..4019
/note="right arm Bb of the major repeat B; putative"
/rpt_family="neurexin repeat"
3709..3735
/note="spliceable sequence, splice site 3"
/evidence=experimental
repeat_region
4146..5225
/note="major repeat C; putative"
/rpt_family="neurexin repeat"
4146..4541
/note="left arm Ca of the major repeat C; putative"
/rpt_family="neurexin repeat"
4599..4718
/note="EGF repeat 3; epidermal growth factor; putative"
/rpt_family="EGF"
4815..5225
/note="right arm Cb of the major repeat C; putative"
/rpt_family="neurexin repeat"
5080..5169
/note="spliceable sequence, splice site 4"
/evidence=experimental
5569..5577
/note="spliceable sequence, splice site 5"
/evidence=experimental
repeat_unit
replace(5569,"")
/note="a" in variation"
variation
BASE COUNT 1680 a 1660 g 1411 t
ORIGIN
Query Match 6.7%; Score 17; DB 3; Length 6349;
Best Local Similarity 100.0%; Pred. NO. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 158 AATGAGCTGTCCTC 174
DB 2348 AATGAGCTGTCCTC 2364

```

Search completed: August 6, 1999, 11:37:13  
Job time: 6319 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:02 ; Search time 3075.15 Seconds  
(without alignments)  
162.285 Million cell updates/sec

Title: US-09-049-696-8

Perfect score: 253  
Sequence: 1 AACAAAGTGTGTCATCATC.....GGCAGAGTGTGAGACAG 253

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	96.4	38.1	455	39	AA871197	AA871197 vq3c11.r
2	81	32.0	435	33	AA443258	AA443258 aa14d09.r
3	73.6	29.1	468	37	AA726662	AA726662 ua93f05.r
4	56	22.1	207	41	AF071903	AF071903 AF071903
5	49.8	19.7	430	47	AI493356	AI493356 tg70t11.x
6	49	19.4	337	50	AV047614	AV047614 AV047614
7	36.4	14.4	587	44	AV044136	AV044136 AV044136
8	31.8	12.6	221	32	AA367484	AA367484 EST78591
9	31.8	12.6	691	49	AA057081	AA057081 AA057081
10	31.4	12.4	440	26	W45744	W45744 mc79e12.r1
11	31.2	12.3	267	20	T29364	T29364 EST78583 Hu
12	31.2	12.3	184	32	AA367525	AA367525 EST78611
13	31.2	12.3	389	34	AA457944	AA457944 v174e09.r
14	31.2	12.3	408	47	AI481293	AI481293 v174e09.x
15	31.2	12.3	448	48	AI552284	AI552284 v174e09.y
16	31	12.3	506	24	H91905	H91905 y881f08.r1
17	30.6	12.1	209	32	AA367483	AA367483 EST78590
18	30.6	12.1	182	32	AA367519	AA367519 EST78604
19	30.6	12.1	771	41	AI007148	AI007148 ua81f12.r
20	30.2	11.9	371	25	N88412	N88412 K3341P Huma
21	30.2	11.9	529	37	AA692991	AA692991 v160e09.s
22	30.2	11.9	473	44	AI262526	AI262526 qk42b05.x
23	30.2	11.9	523	45	AI346357	AI346357 qp50d01.x
24	30.2	11.9	266	48	AI616242	AI616242 v160e09.x
25	30	11.9	471	34	AA495209	AA495209 fa04c08.r
26	30	11.9	737	39	AA882208	AA882208 vx42b04.r
27	30	11.9	473	46	AI437409	AI437409 fb33c09.x
28	29.8	11.8	346	29	AA158121	AA158121 z043g07.s
29	29.8	11.8	537	43	AI162364	AI162364 A016P45U
30	29.6	11.7	290	20	D22278	D22278 R1C10671A
31	29.6	11.7	327	21	D47539	D47539 R1C13108A
32	29.6	11.7	431	28	AA079729	AA079729 zm20e10.r
33	29.6	11.7	314	32	AA336330	AA336330 EST741055
34	29.6	11.7	910	43	AI227295	AI227295 u108a12.y
35	29.6	11.7	526	48	AI587742	AI587742 mp33b08.y
36	29.6	11.7	422	48	AI597604	AI597604 tn15e07.x
37	29.4	11.6	615	24	H67378	H67378 yu52d06..s1
38	29.4	11.6	615	24	H76492	H76492 18197 Lambd
39	29.4	11.6	518	25	N96402	N96402 20981 CD4-1
40	29.4	11.6	194	30	AA222645	AA222645 mv71c03.r
41	29.4	11.6	500	40	AA965347	AA965347 e9c05a1.r
42	29.4	11.6	250	49	AV007538	AV007538 AV007538
43	29.4	11.6	969	26	W27542	W27542 32c1 Human
44	29.2	11.5	493	38	AA808044	AA808044 oc38g06.s
45	29.2	11.5	448	43	AI103566	AI103566 EST712855

#### ALIGNMENTS

RESULT 1  
AA871197 455 bp mRNA EST 16-MAR-1998  
LOCUS vq3c11.r1 Barstead bowel MRLB9 Mus musculus CDNA clone  
DEFINITION IMAGE:1095956 5' similar to SW:ECIC.BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.  
ACCESSION AA871197  
NID 92966642

VERSION	AA871197.1	GI:2966642
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 455) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Gessel,S., Kuaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,K., Tan,F., Underwood,K., Woots,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.	
TITLE	The WashU-HHMI Mouse EST Project	
JOURNAL	Unpublished (1996)	
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1395394.	

Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through ILTL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:602188  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 416.  
 Location/Qualifiers  
     1..455

```

BASE COUNT      122 a      123 c      105 g      105 t
ORIGIN
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1095956"
/clone_lib="Barstead bowel MPLB9"
/tissue_type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTATACCAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCGCTTG]. digested with Not I and cloned into the
Not I and Eco RI sites of the modified pRT73D vector.
Source irradiated bowel harvested 72 hours after
irradiation "(1400 Gys). Library constructed by Bob
Barstead."

```

Query Match	38.1%	Score 96.4	DB 39	Length 455
Best Local Similarity	81.2%	Pred. No. 2,7e-21		
Matches 112	Conservative	0	Mismatches 26	Indels 0
			Gaps	0
QY	116	GAACAAATGGCCCTCATGTAGTCCTTTTGGGCGCCCTTCATCAGGAAATGGAGCTGTCTCTCA	175	
Db	4	GAACAAATGGTCTTTTGTGTATGCTTTGCGACACTCTCCACAGAAATGGCGGATGCTCTCA	63	
QY	176	GCGCTCCATCCAGCTTGAGTAAGGATTAACCTCCAGAAACGCCGCTGGATGAATATGG	235	
Db	64	GCACTTCATCCAGCTGGAGAGCAGGAGGAGTAATCTCCAGAAATTAACCAATGATGAATATGG	123	
QY	236	CACAGTATCGTGGACAG	253	
Db	124	CTCAGTATCGTGGACAG	141	

RESULT 2

AA443258

AA443258 435 bp mRNA EST 03-JUN-1997

DEFINITION *Sal1409.1* Soares\_NhNMPu.S1 Homo sapiens CDNA clone IMAGE:813233 5'

ACCESSION similar to FR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL  
 ; mRNA sequence.  
 AA443258  
 NID G2155933  
 VERSION AA443258.1 GI:2155933  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 435)  
 AUTHORS Hillier,T., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Kucab,T., Lacy,M., Le,N., Lennon,G., Maria,M., Martin,J.,  
 Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
 White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 TITLE Washu-Merck EST Project 1997  
 JOURNAL Unpublished (1997)

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewartson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: -28m3 rev2. ET from Amersham  
 High quality sequence stop: 412.  
 Location/Qualifiers  
     1..435

```

/organism="Homo sapiens"
/db_xref="GDB:6044143"
/db_xref="taxon:9606"
/clone_image="813233"
/clone_id="Soares.NhMPU_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10p"
/notes="Organ: mixed (see below); Vector: pRTT3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2N8M, pregnant uterus
NbHPV, and fetal heart NbHH19w) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

```

Query Match	32.0%	Score 81	DB 33	Length 435
Best Local Similarity	60.8%	Prod. No. 2.7e-16		
Matches 137	Conservative 0	Mismatches 85	Indels 0	Gaps 0
QY	36	GGGCGCCTCTGACGCTCAAGACTAGAGAGAGCTGTCCAAAATGACAGAGAGTTTACAGACA	95	
DB	1	GGTTCAATCTTCACACCCCAAAATCTGAGAGAAATATATACAGCTCTACAGAGAGTTTAAAGTTTC	60	
QY	96	TATGCTTCAGATCAAGTTTGCAGAAATAGGCGCTATGATATCTTTTGGGGCCCTTTATCA	155	
DB	61	TTTGTTCCAAATATATCAAAATCTCCAAATAGATGATGATCTTTCAAGTATAGAAATTTCTCT	120	
QY	156	GGAATGAGAGCTGTCTCTCCAGCGCTCCATCCAGCTTGAGAGATTAAGGATTAAACCTTCAG	215	
DB	121	GGAACTGAGACATTTTCCAGCAACAATATTCAGCTTGAAGGTACAGGTGAATAATGTCAAA	180	
QY	216	AACAGCCAGTGAATGAAATGGCACAGTATCGTGACA	252	
DB	181	CCTCACATCAATTGAAAAACACAGAGACTGTGGATA	217	

RESULT	3	468 bp	MRNA	EST	02-JAN-1998
AA726662					
LOCUS					
DEFINITION	AA726662.1	Stratagene mouse skin (#937313)	Mus musculus cDNA clone		
		IMAGE:1209729.5	similar to SW:ECJC_BOVIN P54281 EPITHELIAL		
ACCESSION	AA726662	CHLORIDE CHANNEL PROTEIN	;	MRNA sequence.	
NID	92744369				
VERSION	AA726662.1	GI:2744369			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 468)				
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,				
	Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,				
	Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,				
	Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and				
	Waston,R.				
TITLE	The Mashu-HHMI Mouse Est Project				
JOURNAL	Unpublished (1996)				
COMMENT	On May 5, 1995 this sequence version replaced gi:798073.				
	Contact: Marra M/Mouse Est Project				
	Mashu-HHMI Mouse Est Project				
	Washington University School of Medicine				
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108				
	Tel: 314 286 1800				
	Fax: 314 286 1810				
	Email: mouseest@watson.wustl.edu				
	This clone is available royalty-free through LNL ; contact the				
	IMAGE Consortium (info@image.llnl.gov) for further information.				
	MG:646073				
	Seq primer: -28m13 rev1 ET from Amersham				
	High quality sequence stop: 443.				
FEATURES	Location/Qualifiers				
source	1..468				
	/organism="Mus musculus"				
	/strain="C57BL/6"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:1209729"				
	/clone_1lb="Stratagene mouse skin (#937313)"				
	/sex="females"				
	/tissue_type="whole skin"				
	/dev_stage="11 weeks old"				
	/lab_host="SOLR (kanamycin resistant)"				
	/note="Organ: skin; Vector: pBluescript SK-; Site_1:				
	Ecobi; Site_2: XhoI; Cloned unidirectionally. Primer:				
	0150 dt. Whole skin from 11 week old C57BL/6 female muce				
	Average insert size: 1.0 kb; Uni-ZAP XR vector; -5'				
	adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor				
	sequence: 5' CTCGAGCTTTTATTTT 3'				
BASE COUNT	121 a 107 c 120 g 120 t				
ORIGIN					
	Query Match	29.1%;	Score 73.6;	DB 37;	Length 468;
	Best Local Similarity	57.3%;	Pred. No. 7.2e-14;		
	Matches 133;	Conservative	0;	Mismatches 19;	Indels 0;
				Gaps	0
OY	4	AAAGGTCGTCATCCACACAGTCGCTTGGGGCCCTCTGAGCTCAAGACTGAGG	63		
		1 1111 111 111 111 111 111 111 111 111 111 111 111 111			
DB	107	ACAGTGATCCACCTTCACTCATTGAGCCCTGGCTTCCTCTGAGCCAGAAAAGTGGGG	156		
OY	64	AGCTTCGCAAAATGACAGAGGTTTACAGACATATCTTCATCAAGTTACAGACATG	123		
		1 111 111 111 111 111 111 111 111 111 111 111 111 111			
DB	167	AATATTCACAGCTTACAGAGAGGTCTAAAGTCTTCATTCAGATAATTTACTTTAATG	226		
OY	124	GCCTATATGATCTTTTGGGGCCCTTTCATCATCAGAAATGAGCTGTCTCTACAGCTCA	183		
		1 111 111 111 111 111 111 111 111 111 111 111 111 111			
DB	227	GAATACGTAACCTTTCGTTGCAATCTCTTCGGAACAGAGACATTTTCACAGCAAGCT	286		

QY	184	TCGACCTTGAGAGTGAAGGATTAACCCCTCCAGACAGCAGGATGATG	235
Db	287	TACAGGTGAGAGCCTGTCCGAAATGTGCAACCCACCCAGCAGCTGCTG	338

  

RESULT	4	AF071903	207 bp	EST	30-JUN-1998
LOCUS	AF071903/c				
DEFINITION	AF071903	Fetal aorta endothelial cell line CPA47		Bos taurus	CDNA
ACCESSION	AF071903	clone 38, mRNA sequence.			
MID	g3265169				
VERSION	AF071903.1	GI:3265169			
KEYWORDS	EST.				
SOURCE	Bos taurus.				
ORGANISM	Bos taurus				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Bovine; Bos.				
REFERENCE	1 (bases 1 to 207)				
AUTHORS	Oberts, T. B. and Allegro, M. C.				
TITLE	Plipen-Binding mRNA				
JOURNAL	Unpublished (1998)				
COMMENT	On May 18, 1995 this sequence version replaced gi.811149.				

  

FEATURES	source
1..207	Location/Qualifiers
/organism="Bos taurus"	
/db_xref="taxon:9913"	
/map="1"	
/clone="38"	
/clone_id="Fetal aorta endothelial cell line CPA47"	
/tissue_type="aorta"	
/cell_type="endothelium"	
/cell_line="CPA47"	
/dev_stage="fetus"	
/note="Plipen(GenBank Accession Number U26024) -Binding mRNA"	

  

BASE COUNT	52 a	46 c	47 g	62 t
ORIGIN				

  

Query Match	22.1%	Score 56;	DB 41;	Length 207;
Best Local Similarity	59.9%;	Pred. No. 2.8e-08;		
Matches 115;	Conservative 0;	Mismatches 70;	Indels 7;	Gaps 1

  

QY	16	TCATCCACACAGTGGCTTTGGGGCCCTTCGACGCTCAGACAACTAGAGAGCTGCCAAA	75
Db	203	TCTAGCCCGCGAGACCTGTGGACCCCTGCGCCGCAAGAACTGAGACATTTGCAATA	144

  

QY	76	TGACAGAGGATTACAGACATATGCTTCAGATCAAGTTCAGAACATGGCCCTCATGTATG	135
Db	143	TGACAGAGGAGATATCGTTTTTTTGGCAATAAAG-----ACATACTGGCCTTACTACG	91

  

QY	136	CTTTGGGGCCCTTCATCAGAGAAATGAGAGCTGTCTCAGCGCTCCATCCAGCTTGAGA	195
Db	90	CTTTCAGAGAGATTTCACTCTAGAACTGGAAGCATCATCAGCAGGCTATTACAGTTGANA	31

  

QY	196	GTAAGGATTA	207
Db	30	GCAAGCCTGA	19

  

RESULT	5	A1493356	430 bp	mRNA	EST	30-MAR-1999
LOCUS	A1493356/c					
DEFINITION	t970f11.x1	Soares_NHMPU.S1	Homo sapiens	CDNA	clone IMAGE:2114157	







The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: [arkerlav@tigr.org](mailto:arkerlav@tigr.org)  
For clone availability, additional sequence and expression

```
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
```

```

map="1, 14: 1, 14"
/clone="IMAGE:849544"
/clone.lib="Soares mouse mammary gland NDMC"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dt) primer [5',
TGTATCCATCATCTGGAAGGAGCGCGCCGAGATGTTTTTTTTTTTTTTT
T 3'] : double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```

	Query Match	Best Local Similarity	Matches 63:	Conservative	Score 31.2%	DB 34	Length 389
					Pred. No. 4.2	Mismatches 53	Indels 0
						Gaps 0	
QY	130	TTGATGCTTTGGGGCCCTTCATGAGAAAGAGAGCTCTCTACGCGCTCCATCCACG	189				
Db	256	TTGATCTCATGTGAGAGCATTTCTCTACAGAGCTCTCTCTGTATGATCAAGCTGAG	197				
QY	190	TTGAGAGTAAAGGAGTTAAACCTCCGAGAAAGCCAGCTGATGATGACACAGTGATC	245				
Db	196	TCGATTTGACACACAAACACGACAGTACAGGCTTGATCTGTGAGCCACCAATGAGAC	141				

RESULT 14  
AI481293

LOCUS	AF1481293	408 bp	EST	09-MAR-1999
DEFINITION	Vf14e09_x1 Soares mouse mammary gland	BDMMG	Mus musculus	CDNA clone
IMAGE	849544 3', mRNA sequence.			

ACCESSION	A1481293
NID	q4374519
VERSION	A1481293.1
REVISION	GI:4374519

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE  
AUTHORS  
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 408)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.

**TITLE** Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Waller, T., Gibbons, M., Page, D., Harvey, N., Schuck, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, T., Cardenas, M., McCann, R.,  
Waterson, R. and Wilson, R.  
**JOURNAL** The WashU-NCI Mouse EST Project 1999  
**COMMENT** On Jun 22, 1998 this sequence version replaced d13247231.  
Unpublished (1999)

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
NCT 501606

This clone was previously sequenced on the 5' end only, this new data is from the 3' end  
High quality sequence stop: 351.

```

FEATURES
source
Location/Qualifiers
1. .408
/organism="Mus musculus"

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:848544"
/clone_lib="Soares mouse mammary gland NbMxG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3p-Pac
Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTACCAATCTGAAGGAGGAGCGCGCCGCAATGTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```

	Query Match	Best Local Similarity	Matches	Score 31.2	DB 47	Length 408	Pred. No. 4.3	Mismatches 53	Indels 0	Gaps 0
Qy	130	TTGATGCTTTGGGGCCCTTTCATCAGGAATGAGAGCTGTCTCTAGCGCTCATCCAGC	189							
Db	273	TTGATCTCATGAGAGCGCATTTCTCTCACTGAGAGCTCTTCTCTGATTAATCCACTGTAG	332							
Qy	190	TTGAGAGTAAGGAGATTAAACCTCCAGAAACAGCCAGTGATGATGACAGATGATC	245							
Db	333	TCAAGTTGACACACAAAACAGCCAGTACAGCCCTGGATCTGTGAGACCCCAATGAGC	388							

RESULT 15

LOCUS	448 bp	mRNA	EST	23-MAR-1999
AI552284				
VI74699.y1	Soares mouse mammary gland	NbMmg	Mus musculus cDNA clone	
IMAGE:849544	5', mRNA sequence.			

ACCESSION	AI552284
NID	94484647
VERSION	AI552284.1
	GI:4484647

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE  
AUTHORS

Eukaryota: Metazoa; Chordata; Cranata; Vertebrata; Mammalia  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1. (bases 1 to 448)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, J.

**TITLE** Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
**JOURNAL** Gibson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schirk, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McClann, R.,  
 Waterson, R., and Wilson, R.  
**COMMENT** The WashU-NCI Mouse EST Project 1999  
 Unpublished (1999)  
 On Mar 10, 1998 this sequence version replaced 41:2948905.

Contact: Marra M/Mashu-NCI Mouse ESR Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 USA

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)

This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
MGI:501696

This read is a RESEQUENCE of a previously sequenced mouse clone. This read has been verified (found to hit its original self in the correct orientation).  
Seq primer: -40RP from Glibco  
High quality sequence stop: 447.



## FEATURES

**Source**

Location/Qualifiers  
1. .448

```
/organism="Mus musculus"  
/strain="C57BL/6J"
```

```
/db_xref="taxon:10090"
/map="12"
```

```

/map="12"
/clone="IMAGE:849544"
/cjone_1jh="Scars mouse mammary a)and NbMNG"

```

```
/clone_lib="soares mouse man"
/sex="male"
```

```

/tissue_type="mammary gland"
/dev_stage="4 weeks"

```

```

/uev_stage="4 weeks
/lab_host="DH10B"
/note="Organ: mammary gland"

```

```
/note="Organ: mammary gland,  
(Pharmacia) with a modified  
site 2: Eco RI: 1st strand
```

Site\_2: Eco RI; 1st strand cDNA was primed with oligo(dT) primer [5',

TGTTACCAATCTGAAGTGGAGCGGCCGAATGTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI

1.5 µl, double stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector.

the Not I and Eco RI sites of the modified pI/T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M Fatima

constructed and normalized by Bento Soares and M. Fatima Bonaldo."

a	103 c	106 g	117 t
---	-------	-------	-------

```
12.38; Score 31.2; DB 48; Length 448;
5438: Pred No 4 4:
```

ity 54.38; Pred. No. 4.4;  
servative 0; Mismatches 53; Indels 0; Gaps 0;

TTGGGCCCCCTTCATCAGGAATGGAGCTGTCTCTCAGCGCTCCATCCAGC 189

TGGAGGCAATTCCTCAACTGAAGCTTCTCTGATAACTCCAGCTGAG 214  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

TGGAGGAGGCAATTCCCTCAACCTGAAGCTCCCTTCTCTGATAACTCCAGCGTAG 214  
 TGGGATTAACCCCTCCAGAACACAGTGGATGAATGAGCCACTAGTATC 245

55GGATTAAACCCCTCCAGAACAGCCAGTGATGTAATGGCACAGTGA TC 245

ACACAAACCAGCCAGTACAGCCTGGATCTGTGAGCCACAATGAGC 158

Aug 6, 1999, 13:33:04

```
Search completed: August 6, 1999, 13:33:04
Job time: 12982 sec
```

Job time: 12982 sec

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:11 ; Search time 3075.15 Seconds  
(without alignments)  
122.516 Million cell updates/sec

Title: US-09-049-696-11

Perfect score: 191  
Sequence: 1 GGCTTGTAGTGGACAAAAA.....AACAGGACACGCAATT 191

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST.\*

1: em\_est1:.\*  
2: em\_est2:.\*  
3: em\_est3:.\*  
4: em\_est4:.\*  
5: em\_est5:.\*  
6: em\_est6:.\*  
7: em\_est7:.\*  
8: em\_est8:.\*  
9: em\_est9:.\*  
10: em\_est10:.\*  
11: em\_est11:.\*  
12: em\_est12:.\*  
13: em\_est13:.\*  
14: em\_est14:.\*  
15: em\_est15:.\*  
16: em\_est16:.\*  
17: em\_est17:.\*  
18: em\_est18:.\*  
19: em\_est19:.\*  
20: gb\_est1:.\*  
21: gb\_est2:.\*  
22: gb\_est3:.\*  
23: gb\_est4:.\*  
24: gb\_est5:.\*  
25: gb\_est6:.\*  
26: gb\_est7:.\*  
27: gb\_est8:.\*  
28: gb\_est9:.\*  
29: gb\_est10:.\*  
30: gb\_est11:.\*  
31: gb\_est12:.\*  
32: gb\_est13:.\*  
33: gb\_est14:.\*  
34: gb\_est15:.\*  
35: gb\_est16:.\*  
36: gb\_est17:.\*  
37: gb\_est18:.\*  
38: gb\_est19:.\*  
39: gb\_est20:.\*  
40: gb\_est21:.\*  
41: gb\_est22:.\*  
42: gb\_est23:.\*  
43: gb\_est24:.\*  
44: gb\_est25:.\*  
45: gb\_est26:.\*  
46: gb\_est27:.\*  
47: gb\_est28:.\*  
48: gb\_est29:.\*  
49: gb\_est30:.\*  
50: gb\_est31:.\*  
51: gb\_est32:.\*  
52: em\_est20:.\*  
53: em\_est21:.\*

54: em\_est22:.\*  
55: em\_est23:.\*  
56: em\_est24:.\*  
57: em\_est25:.\*  
58: em\_est26:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	123.8	64.8	455	39	AA871197	AA871197 vg32c11.r
2	105.6	55.3	479	35	AA541829	AA541829 vj01f01.r
3	65	34.0	450	37	AA734088	AA734088 vs19c08.r
4	32.4	17.0	502	38	AA57956	AA57956 zq49g01.s
5	32.2	16.9	386	21	T78952	T78952 yq21a08.r1
6	31.2	16.3	442	27	AA009744	AA009744 zeb2g01.s
7	31.2	16.3	432	42	A1082401	A1082401 ov21c10.x
8	31.2	16.3	449	42	A1140962	A1140962 qf24f01.x
9	31.2	16.3	414	45	A1361578	A1361578 qv48e11.x
10	31.2	16.3	432	46	A1421702	A1421702 tf54d09.x
11	31.2	16.3	472	46	A1432456	A1432456 tg73g06.x
12	31.2	16.3	440	47	A1524059	A1524059 th01c03.x
13	31.2	16.2	496	22	R44318	R44318 yq35b09.s1
14	30.8	16.1	530	24	H76962	H76962 17393 Lambd
15	30.8	16.1	490	25	N85495	N85495 j3752E Huma
16	30.8	16.1	331	25	N90877	N90877 zbl1c10.s1
17	30.4	15.9	360	36	C65063	C65063 yvj1
18	30.2	15.8	420	23	H51345	H51345 y030d05.r1
19	30	15.7	451	22	R42834	R42834 yq05a09.s1
20	30	15.7	444	25	N66205	N66205 yv67g03.s1
21	29.8	15.6	486	23	H18471	H18471 ym43h05.s1
22	29.8	15.5	412	44	A1278076	A1278076 qm57g05.x
23	29.6	15.5	321	20	Z41208	Z41208 HSC2RC012.n
24	29.6	15.5	438	21	R07193	R07193 y114f01.s1
25	29.6	15.5	309	21	R07194	R07194 y114f02.s1
26	29.6	15.5	470	21	R13627	R13627 yf60b04.r1
27	29.6	15.5	292	36	AA634820	AA634820 ab28d07.r
28	29.4	15.4	193	33	AA433368	AA433368 TEMP0230
29	29.2	15.3	415	49	AV003184	AV003184 AV003184
30	29	15.2	400	23	H30430	H30430 ym58a12.r1
31	29	15.2	451	27	W82768	W82768 m103b08.r1
32	29	15.2	412	28	AA068843	AA068843 mm59g11.r
33	29	15.2	234	34	AA474924	AA474924 vth08d08.r
34	29	15.2	413	45	A1325676	A1325676 mm59g11.y
35	29	15.2	680	47	A1527737	A1527737 u128e09.y
36	28.8	15.1	397	22	R61457	R61457 yhl5b12.s1
37	28.6	15.0	407	23	F02709	F02709 HSC17F062.n
38	28.6	15.0	318	20	H25999	H25999 y151e10.s1
39	28.6	15.0	375	28	AA121019	AA121019 zn84e11.r
40	28.6	15.0	547	29	AA166705	AA166705 z085b07.s
41	28.6	15.0	450	37	AA687377	AA687377 nv62a03.s
42	28.4	14.9	548	40	AA944049	AA944049 EST119548
43	28.4	14.9	397	41	A1043992	A1043992 UI-R-CI-J
44	28.4	14.9	286	49	AV024450	AV024450 AV024450
45	28.2	14.8	478	40	AA988739	AA988739 os15b02.s

#### ALIGNMENTS

RESULT 1  
LOCUS AA871197 455 bp mRNA EST 16-MAR-1998  
DEFINITION vg32c11.r1 Barstead bowel MRLB9 Mus musculus CDNA clone  
IMAGE:1095956 5' similar to SW:ECIC\_BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.  
ACCESSION AA871197  
NID g2966642

VERSION	AA871197.1	GI:29666542
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Weising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.	
TITLE	The Mashu-HMI Mouse EST Project	
JOURNAL	Unpublished (1996)	
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1395394.	

Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through INTL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:602188  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 414.  
 Location/Qualifiers  
 1..455

BASE COUNT	122 a	123 c	105 g	105 t
ORIGIN	/organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:1095956" /clone_lib="Barstead bowel MPLRB9" /tissue_type="bowel" /dev_stage="8 weeks" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dn) primer [5', TGTTACGATCTGTAAGTGGGCGCGCCCTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGATCTCG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Source irradiated bowel harvested 72 hours after irradiation (1400 Gys). Library constructed by Bob Barstead."			
Query Match	64.8%	Score 123.8;	DB 39;	Length 455;
Best Local Similarity	78.0%	Pred. No. 1.1e-30;		
Matches 149; Conservative	0;	Mismatches 42;	Indels 0;	Gaps 0;

Query Match	64.8%	Score 123.8	DB 39	Length 455
Best Local Similarity	78.0%	Pred. No. 1.1e-30		
Matches 149	Conservative	0	Mismatches 42	Indels 0
QY	1	GGCTTGTAGTGACAAAAACACCAAAATGCGCTACTCTCAATTCCAGGACATTGCTAAG	60	
Db	233	GGTTTATTACTAACAACAACCACTAAGGTGGCTACTCCAAGTCCAGGACGGCTAAG	292	
QY	61	GTTGGCACTGGAAATACAGTCTGCAAGCAACCTCACAACCTTGACCCCTGACTGTACG	120	
Db	293	GTTGGCTTTTGGAAATACAGCAATTCAAGGAGCTACAGACTCTCACCCTTGACCTGAC	352	
QY	121	TCCCGTGCTCAATGCTACCTGCTCTCAATTACAGTACTTCAAAAGCAAGCAAGAGAC	180	
Db	353	TCCCGTGCAAGCAAGTCTCACTGCCTCCTATTACAGTACCCCGGTAGTAATTAAGAAC	412	
QY	181	ACCAACAATT 191		
Db	413	ACAGGAAATT 423		

RESULT	2
AA541829	
LOCUS	
DEFINITION	AA541829 479 bp mRNA EST 30-UTL-1997 v 01c01.r1 Batstad mouse pooled organs MRLR4 Mus musculus cDNA clone IMAGE:920473 similar to TR.G1184066 G1184066
ACCESSION	CALCULUM-ACTIVATED CHLORIDE CHANNEL. ; mRNA sequence.
NID	AA541829
VERSION	92288263
KEYWORDS	AA541829.1 GI:2288263
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 479)
AUTHORS	Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisell,S., Kucab,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Martinwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R.
TITLE	The WashU-HMNI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1406841.

TITLE	The WashU-HMMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:11406841.
FEATURES	Contact: Marra M/Mouse EST Project WashU-HMMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LINT; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:532689 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 368. Location/Qualifiers 1..479
SOURCE	

```

source
1. 479
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image:920473"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5
TGTTCGAATCCTGAATGGAGCGCGCCGCTTTTCTTTTCTTTTCTTTT
3)). double-stranded cDNA was ligated to Eco RI adaptors
(CTGTGATTCGGTACCT), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library constructed by Bob Barstead."
139 a 115 c 120 g 105 t

```

	Query Match	Score	DB	Length
	Best Local Similarity	78.8%	Pred. No.	1.e-24;
	Matches 126;	Conservative	0;	Mismatches 34; Indels 0; Gaps 0;
OY	32	CCACACTCCAATTCACAGGCACTTGTAAAGTTGGACATTGGAAATACAGCTTCAAAGCAA	91	
Db	5	CCAACTCCCAAGTCCCACAGCACGGCTAAGTGTCITTTGGAAATACAGCATCAAGCA	64	
OY	92	GCCACAACCTTGACCTGACTGTCAAGSTGCCGTGCATAGCTACCCTGCTCCAA	151	
Db	65	GCCACAAGACTCTCAGCTTGACTGTCACTCTCCCTCGACGAATGCTCACTGCTCTCTTA	124	
OY	152	TTCAGTGACTTCCAAAAGCAACAAGGACACAGCAATT	191	

Db 125 TTACAGTGACCCCGGTAGTGAATAGACACAGGAATT 164

RESULT 3  
AA734088 450 bp mRNA EST 07-JAN-1998  
LOCUS AA734088  
DEFINITION vs19c08.r1 Barstead clone irradiated colon MPLRB7 Mus musculus cDNA  
clone IMAGE:1138670 5' similar to SW:BCLC\_BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.

ACCESSION AA734088  
NID 92755755  
VERSION AA734088.1 GI:2755755  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 450)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Stepcoe, M., Tan, F., Underwood, R., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
COMMENT On May 8, 1995 this sequence version replaced gi:801722.

TITLE  
JOURNAL  
COMMENT

CONTACT: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:619942  
Possible reversed clone: similarity on wrong strand  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 444.  
Location/Qualifiers  
1..450  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone\_lib="IMAGE:1138670"  
/clone\_lib="Barstead mouse irradiated colon MPLRB7"  
/dev\_stage="8 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained  
from 8 week old mouse. Colon was harvested 72 hours after  
irradiation with 1400 Gys. 1st strand cDNA was primed  
with a Not I - oligo(dT) primer  
15'TGTTACGATCTGACGTGAGCGGCCGCCCTTTTCTTTTCTTTTCTTTTCTTTT  
T3'; double-stranded cDNA was ligated to Eco RI  
adaptors (AATTCGATCTTG), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT773  
vector. Library constructed by Bob Barstead."

BASE COUNT 132 a 109 c 112 g 97 t

ORIGIN

Query Match 34.0%; Score 65; DB 37; Length 450;  
Best Local Similarity 76.2%; Pred. No. 2.2e-11;  
Matches 80; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 87 AGCAAGCTCAACACCTTGACCTGACGTGACGTCCCGCGCTCCATGCTACCTGCC 146  
|||||  
DB 12 AGGAGCTCAACACCTTGACCTGACGTGACGTCCCGCGCTCCATGCTACCTGCC 71  
|||||

QY 147 TCAATTACAGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGT 191  
|||||

Db 72 TCCTATTACAGTGACCCCGGTAGTGAATAGACACAGGAATT 116

RESULT 4  
AA757956 502 bp mRNA EST 23-JAN-1998  
LOCUS AA757956/c  
DEFINITION z649601.s1 Soares pineal gland N3HPG Homo sapiens cDNA clone  
IMAGE:396720 3' similar to gb:104456 GALECTIN-1 (HUMAN); mRNA  
sequence.

ACCESSION AA757956  
NID 92805819  
VERSION AA757956.1 GI:2805819  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 502)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Kritman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Stepcoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2153478.

TITLE  
JOURNAL  
COMMENT

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 195.  
Location/Qualifiers  
1..502  
/organism="Homo sapiens"  
/db\_xref="GDB:130262"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:396720"  
/clone\_lib="Soares\_pineal\_gland\_N3HPG"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: pineal gland; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
15' TGTTACGATCTGACGTGAGCGGCCGCCCTTTTCTTTTCTTTTCTTTTCTTTT  
3'); double-stranded cDNA was size selected, ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified pT73  
vector (Pharmacia). Library constructed by Bento Soares  
and M.Fatima Bonaldo."

BASE COUNT 94 a 127 c 148 g 133 t

ORIGIN

Query Match 17.0%; Score 32.4; DB 38; Length 502;  
Best Local Similarity 60.0%; Pred. No. 1.1;  
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 99 AACCTGACCCCTGACGTGACGTCCCGGTGCTCAATGCTACCTCCCTCAATTACAGT 158  
|||||  
DB 361 AACCTGACCCCTGACGTGACGTCCCGGTGCTCAATGCTACCTCCCTCAAGCCAGC 302  
|||||

QY 159 GACTTCCAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 188  
|||||

DB 301 AGACGCCAACACCATGATGATGACAGCA 272

RESULT 5  
T78952/c  
LOCUS T78952 386 bp mRNA EST 15-MAR-1995

[illegible]

```

ACCESSION      IMAGE:365520 3', mRNA sequence.
NID            AA009744
VERSION        51470547
KEYWORDS       AA009744.1  GI:1470547
SOURCE         EST.
ORGANISM       human.
               Homo sapiens
REFERENCE      Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
               Eutheria: Primates: Catarrhini: Hominoidea: Homo.
               1 (bases 1 to 442)
               Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapellil,B.,
               Chisoe,S., Dietrich,N., Dubugue,T., Favallo,A., Gish,W.,
               Hawkins,M., Hiltman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
               Marids,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
               Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
               Trevischi,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
               and Marra,M.
               Generation and analysis of 280,000 human expressed sequence tags
               Genome Res. 6 (9), 807-828 (1996)
               97044478
TITLE          On Apr 14, 1993 this sequence version replaced gi:693428.
JOURNAL        MEDLINE
COMMENT        Contact: Wilson RK
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: estewatson.wustl.edu
               This clone is available royalty-free through LNL : contact the
               IMAGE Consortium (info@image.lnl.gov) for further information.
               Insert Length: 928 Std Error: 0.00
               Seq primer: mob.REGA+ET
               High quality sequence stop: 316.
               Location/Qualifiers
                   1..442
                   /organism="Homo sapiens"
                   /db_xref="GDB:128224"
                   /db_xref="taxon:9606"
                   /map="1"
                   /clone="IMAGE:365520"
                   /clone_lib="Soares_fetal_heart_NbHH19w"
                   /sex="unknown"
                   /dev_stage="19 weeks"
                   /lab_host="DH10B (ampicillin resistant)"
                   /note="Organ: heart; Vector: pTTT3D (Pharmacia) with a
                   modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                   strand cDNA was primed with a Not I - oligo(dT) primer [5
                   TGTTACCAATCTGAAGTGGAGCGCGCCATCTTTTCTTTTCTTTT 3'],
                   double-stranded cDNA was size selected, ligated to Eco
                   RI adapters (Pharmacia), digested with Not I and cloned into
                   the Not I and Eco RI sites of a modified pTTT3 vector
                   (Pharmacia). Library went through one round of
                   normalization to a Cot = 5. Library constructed by
                   M.Felina Bonaldo. This library was constructed from the
                   same fetus as the fetal lung library, Soares fetal_lung
                   NbHH19w."
BASE COUNT     90 a 108 c 158 g 83 t 3 others
ORIGIN
Query Match    16.3%: Score 31.2; DB 27; Length 442;
Best Local Similarity 57.0%: Pred. No. 2.7;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
0Y 84 GCACGAGCTCACAAACCTTGACCCCTACGTCACGTCGCCGTGGTCCAAATGCTACCT 143
DB 410 GGAGCGCTGAGAGCCCGAGCCCTGCGCCCTTACACACGCTCCCTTGGGTCCAGGGGCTTTTC 351
0Y 144 GCCTCCATTTACAGTGAATCTCCAAACGACAGACACC 183
DB 350 GGCCTCATTTGACAGGCGCTGCCAATGCTTCACGACCCC 311

```

RESULT	7
A1082401/c	
LOCUS	A1082401 432 bp mRNA EST 27-AUG-1998
DEFINITION	ov1c10.x1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1637970 3' , mRNA sequence.
ACCESSION	A1082401
NID	g3419193
VERSION	A1082401.1 GI:3419193
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 432)
REFERENCE	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE	Tumor Gene Index,
JOURNAL	Unpublished (1997)
COMMENT	On Jan 17, 1998 this sequence version replaced gi:2043954.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmerit-Buck, M.D., Ph.D.  
cDNA library preparation: M. Bento Soares, Ph.D.  
cDNA library arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://bio.lnl.gov/dbw/image/image.html>

```

FEATURES
  source
    Insert Length: 764      Std Error: 0.00
    Seq primer: ~40ml3 fwd.  Er from Amersham
    High quality sequence stop: 414.
    location/Qualifiers
      1. .432

```

```

FEATURES
    source
        location/Qualifiers
            1. .432
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1637970"
            /clone_11b="NCI-CGAP_Br2"
            /sex="female, pooled"
            /tissue-type="breast"
            /lab_host="DH10B"
            /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker; 1st strand cDNA was prepared from pooled bulk
            breast tumor tissue, and was then primed with a Not I -
            oligo(dT) primer. Double-stranded cDNA was ligated to Eco
            RI adaptors (Pharmacia), digested with Not I and cloned
            into the Not I and Eco RI sites of the modified pT73
            vector. This library is the normalized version of
            NCI-CGAP_Brl.1. Library was constructed by Bento Soares
            and M. Fatima Bonaldo."
BASE COUNT
    93 a      113 c      163 g      63 t
ORIGIN

```

[illegible]

DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
gq44f01.x1 NC1_CGAP_Brr25 Homo sapiens cDNA clone IMAGE:1750969 3 ,											
mRNA sequence.	A1140962	g3648419	A1140962.1	GI:3648419	EST.	human.					
						Homo sapiens					
						Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
						Eutheria; Primates; Catarrhini; Homnidae; Homo.					
						1 (bases 1 to 449)					
						NC1/NIHDS-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .					
						National Cancer Institute / National Institute of Neurological					
						Disorders and Stroke, Brain Tumor Genome Anatomy Project					
						(CGAP/BrrAP), Tumor Gene Index					
						Unpublished (1998)					
						On Jan 19, 1998 this sequence version replaced gi:2286357.					

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert\\_Strausberg@nih.gov](mailto:Robert_Strausberg@nih.gov)  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.  
cDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
cDNA library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/dbip/image/image.html](http://www-bio.llnl.gov/dbip/image/image.html)

```

FEATURES
  source
    1. .449
    Insert Length: 560   Std Error: 0.00
    Seq primer: -40ml3 fwd.  ET from Amersham
    High quality sequence stop: 412.
    location/Qualifiers

```

```

source
1. .449
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1750969"
/clone_1ib="NCI-CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: p7R3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
stranded cDNA was primed with a Not I - oligo(dT) primer [5'
TGATCCACATCTGAAGTGGAGCGCCGACATGTTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p7R3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fátima Bonaldo."
BASE COUNT
100 a 100 a
119 c 165 g 65 t

```

Query Match	16.3%	Score 31.2	DB 42	Length 449
Best Local Similarity	57.0%	Pred. No. 2.7		
Matches 57	Conservative 0	Mismatches 43	Indels 0	Gaps 0
OY	84	GCACGACAGCTCACAAACCTTGACCCCTGACTGTCAGCTCCGCTGCCTCAATGCTACCT	143	
Db	390	GGAGCGCTCGAACCCCGAGCCCTGCCTCCCTACCAAGCTCCCTTGGGGTCCAGGGGCTTTC	331	
OY	144	GCTCTCAATTACAGTGAATCTCCAAACGACAGACGACAC	183	
Db	330	GGCTCATATTTCAGAGGGCTCGCCAAATGCTTACGACACCC	291	
RESULT	9			
LOCUS	AI361578/c			
DEFINITION	AI361578	414 bp	mRNA	EST
	gy48bell.x1 NCI_CGAP_Brn23		Homio sapiens	CDNA clone
	mRNA sequence.			IMAGE:2015276
				3'

ACCESSION A1361578  
NID G4113199  
VERSION A1361578.1 GI:4113199  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Euthera; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 414)  
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
TITLE (CGAP/BTCAP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1405296.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bdrip/image/image.html](http://www-bio.llnl.gov/bdrip/image/image.html)

Insert Length: 932      Std Error:  
Seq primer: -40UP from G1bco  
High quality sequence stop: 374.

## FEATURES

**Source**

BASE COUNT  
ORIGIN

86 a	110 c	157 g	61 t
------	-------	-------	------

Query Match	16.3%	Score 31.2;	DB 45;	Length 414;
Best local Similarity	57.0%	Pred. No.2.6;		
Matches 57;	Conservative 0;	Mismatches 43;	Indels 0;	Gaps 0;

QY	84	GCACGCAAGCTGCACAAACCTTGACCCCTGACGTACGCTCCCGTGCCTGCACATGGTACCCT	143
Db	388	GGAGCGCTGACGCCCCACGACCCCTGGCCCCCTGACACGCTCCCTCCCTTGGTTCACAGGGGTCTTTC	3299
QY	144	GCCCTCATTTCAGTGTACCTTCGAAAACGAAACAGACACCC	183
Db	328	GGCTCATATTGACAGGGCTGCGCATGCTTACACGACACCC	289

RESULT	10
A1421702/c	
LOCUS	A1421702/c
DEFINITION	t15f4d09.x1 NC1_CGAP_Brn23 Homo sapiens mRNA sequence.
ACCESSION	A1421702
ID	94267633

VERSION	AI421702.1	GI:4267633
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
	Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	1 (bases 1 to 432)	
AUTHORS	NCL/NINDS-CGAP <a href="http://www.ncbi.nlm.nih.gov/nclgap">http://www.ncbi.nlm.nih.gov/nclgap</a> .	
TITLE	National Cancer Institute / National Institute of Neurological	
	Disorders and Stroke, Brain Tumor Genome Anatomy Project	
	(CGAP/STGAP), Tumor Gene Index	
JOURNAL	Unpublished (1998)	
COMMENT	on Apr 7, 1998 this sequence version replaced gi:3035239.	

Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
[www-bio.llnl.gov/dbirp/image/image.html](http://www.bio.llnl.gov/dbirp/image/image.html)

Insert Length: 928      Std Error:  
Seq primer: -40UP from Gibco  
High quality sequence stop: 422.

## FEATURES

## Source

BASE COUNT  
ORIGIN

94 a	113 c	164 g	61 t
------	-------	-------	------

Query Match	16.38;	Score 31.2;	DB 46;	Length 432;
Best Local Similarity	57.08;	Pred. No. 2.7;		
Matches 57; Conservative	0;	Mismatches 43;	Indels 0;	Gaps 0

QY	84	GCACAGCAGCTACCAACCTTGACCCCTGATGTGACAGTCCCGCGCTCCATGCTACCGT	143
Db	389	GGACGGCTGAGCCCCAGGCCCGGCCCTGACACACCTCCCTTGGGTCCAGGGGTCTTC	330
QY	144	GCCTCCATTATGACGACTTCCAAAAGCAACGAGCACC	183
Db	329	GGCTCATATTGACGGGCTCGCCATGCTTTCACGACGCCCC	290

RESULT	11				
LOCUS	AI432456/c				
DEFINITION	AI432456	472 bp	mrna	EST	30-MAR-1999
ACCSSION	t973005.x1	Soares	NHMPu_S1	Homo sapiens	CDNA clone IMAGE:2114458
NID	AI432456				
VERSION	94281842				
KEYWORDS	AI432456.1	GI:4281842			
	EST.				



SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 472)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2948964.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Insert Length: 764 Std Error: 0.00  
 Seq primer: -40UP from G1bco  
 High quality sequence stop: 445.  
 FEATURES  
 source  
 1..472  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2114458"  
 /clone\_lib="Soares.NhHMPU.S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and  
 pregnant uterus"  
 /lab\_host="DH10B"  
 /note="Organ: mixed (see below); Vector: p773D-Pac  
 (Pharmacia) with a modified polylinker. Site.1: Not I;  
 Site.2: Eco RI. Equal amounts of plasmid DNA from three  
 normalized libraries (melanocyte 2NBH, pregnant uterus  
 NBHP, and fetal heart NBH19W) were mixed, and ss circles  
 were made in vitro. Following HAP purification, this DNA  
 was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from pools of  
 5,000 clones made from the same 3 libraries. The pools  
 consisted of I.M.A.G.E. clones 260232-265223,  
 340488-345479, and 484488-489479."  
 BASE COUNT 101 a 123 c 166 g 82 t  
 ORIGIN  
 Query Match 16.3%; Score 31.2; DB 46; Length 472;  
 Best Local Similarity 57.0%; Pred. No. 2.8;  
 Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
 QY 84 GCAGCAGCTCAACAACCTTGACCTGACGTGACGTCGCCGTGCAATGCTACCT 143  
 DB 406 GGACGCTGAGCCCCAGCCCTGCCCCCTGACCAAGCTCCCTTGGGTCCAGGGGTCTTTC 347  
 QY 144 GCCTCCATTACAGTACTTCCAAAAGCAAGACACCC 183  
 DB 346 GGCTCATATTGACGAGGCGCTGCCAATGCTTCACGACACCC 307

RESULT 12  
 LOCUS AI524059 440 bp mRNA EST 13-APR-1999  
 DEFINITION th01c03.x1 NCI-CGAP-CLL1 Homo sapiens cDNA clone IMAGE:2116996 3',  
 mRNA sequence.  
 ACCESSION AI524059  
 NID 94438194  
 VERSION AI524059.1 GI:4438194  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 440)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL Unpublished (1997)  
 COMMENT On Apr 7, 1998 this sequence version replaced gi:3034525.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,  
 M.D., Louis M. Staudt, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
 Insert Length: 947 Std Error: 0.00  
 Seq primer: -40UP from G1bco  
 High quality sequence stop: 431.  
 FEATURES  
 source  
 1..440  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="6: 21"  
 /clone="IMAGE:2116996"  
 /clone\_lib="NCI-CGAP-CLL1"  
 /tissue\_type="B-cell, chronic lymphocytic leukemia"  
 /lab\_host="DH10B"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified  
 polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand CDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACATCGAGTGGAGGCGCCGCTCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded CDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified p773 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M. Fatima Bonaudo."  
 BASE COUNT 94 a 113 c 164 g 69 t  
 ORIGIN  
 Query Match 16.3%; Score 31.2; DB 47; Length 440;  
 Best Local Similarity 57.0%; Pred. No. 2.7;  
 Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
 QY 84 GCAGCAGCTCAACAACCTTGACCTGACGTGACGTCGCCGTGCAATGCTACCT 143  
 DB 397 GGACGCTGAGCCCCAGCCCTGCCCCCTGACCAAGCTCCCTTGGGTCCAGGGGTCTTTC 338  
 QY 144 GCCTCCATTACAGTACTTCCAAAAGCAAGACACCC 183  
 DB 337 GGCTCATATTGACGAGGCGCTGCCAATGCTTCACGACACCC 298

RESULT 13  
 LOCUS R44318 496 bp mRNA EST 22-MAY-1995  
 DEFINITION y935b09.s1 Soares Infant brain IN1B Homo sapiens cDNA clone  
 IMAGE:34382 3', mRNA sequence.  
 ACCESSION R44318  
 NID 9821289  
 VERSION R44318.1 GI:821289  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 496)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maria, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)  
COMMENT On May 9, 1995 this sequence version replaced gi:803042.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

Insert Size: 1256  
High quality sequence stops: 321 Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1256 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 321.  
Location/Qualifiers

FEATURES  
SOURCE

1..496  
/organism="Homo sapiens"  
/db\_xref="GDB:406729"  
/db\_xref="taxon:9606"  
/clone="IMAGE:34382"  
/clone\_lib="Soares Infant Brain INTB"  
/sex="female"  
/dev\_stage="73 days post natal"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: whole brain; Vector: Latmid BA; Site:1: Not  
I; Site:2: Hind III; 1st strand cDNA was primed with a Not  
I - oligo(dT) primer [5'  
AAGTGAAGAATTCGGCCGACGAGATTTTCTTTTCTTTT 3'];  
double-stranded cDNA was ligated to Hind III adaptors  
(pharmacia), digested with Not I and directionally cloned  
into the Not I and Hind III sites of the Latmid BA vector.  
Library went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."  
BASE COUNT 149 a 99 c 130 g 109 t 9 others  
ORIGIN

Query Match 16.2%; Score 31; DB 22; Length 496;  
Best Local Similarity 53.5%; Pred. No. 3.3;

Matches 61; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1 GCCTTTAGTGGACAAAACACCAATGGCTACCTCAATCCAGGCAATTGCTAAG 60  
DB 444 GGTGTGCAATAAAAAACCCTCCGNGGTAAACCCCACTCCGGAGAGTCTCGG 385  
QY 61 GTTGCACTTGGAAATACAGTCTGCAGCAAGCTACCAACCTGACCTGACT 114  
DB 384 CTCTGTGCTTAAAGTCCCTCAGCAACCGTGACCTTAATTAACCTTGCT 331

RESULT 14  
LOCUS H76962 530 bp mRNA EST 05-JAN-1998  
DEFINITION 17393 Lambda-PR2 Arabidopsis thaliana cDNA clone 200P2377, mRNA  
sequence.  
ACCESSION H76962  
NID g1054213  
VERSION H76962.1 GI:1054213  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
Arabidopsids.  
REFERENCE 1 (bases 1 to 530)  
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,  
McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasnow,M.,  
Retzel,E. and Somerville,C.  
TITLE Genes galore: a summary of methods for accessing results from  
large-scale partial sequencing of anonymous Arabidopsis cDNA clones

JOURNAL Plant Physiol. 106, 1241-1255 (1994)  
MEDLINE 95148729  
COMMENT On Nov 29, 1993 this sequence version replaced gi:430429.

Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI  
Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 22313c@nslm.cl.msu.edu

Seq primer: T7 dye primer  
Location/Qualifiers

FEATURES  
SOURCE

1..530  
/organism="Arabidopsis thaliana"  
/strain="var columbia"  
/db\_xref="taxon:3702"  
/clone="200F2377"  
/clone\_lib="Lambda-PRL2"  
/note="Vector: lambda zip-lox; Site:1: Sal; Site:2: Not;  
lambda PRL2 is a cDNA library derived from equal  
quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
day germinated etiolated seedlings; 2) tissue culture  
grown roots; 3) staged plants half with 24 hour light  
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
same plants as 3 but aerial tissue (stems, flowers and  
siliques. The vector is BRU's lambda zip-lox. The cDNA  
inserts were directionally cloned with Sal-Not arms using  
oligo dT primed cDNA."  
BASE COUNT 160 a 113 c 126 g 110 t 21 others  
ORIGIN

Query Match 16.1%; Score 30.8; DB 24; Length 530;  
Best Local Similarity 56.0%; Pred. No. 3.9;

Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 48 AGCATTCCTAAGTGGCACTTGGAAATACAGTCTGCAGCAAGCTACCAACCTGAC 107  
DB 343 ANGCAATTCCTTCCTCAAGATGTGGAAATATGGCAAGTCAATCTACACCTGAA 284  
QY 108 CCTGACTGTCAAGTCCGCTGGTCCCAATGCATACCTGGCT 147  
DB 283 CTGCTCATCAGTGACCTTCATCTCAACCTTACT 244

RESULT 15  
LOCUS N85495 331 bp mRNA EST 01-APR-1996  
DEFINITION J3752F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA  
clone J3752 5' similar to BETA-GALACTOSIDE-BINDING LECTIN, mRNA  
sequence.  
ACCESSION N85495  
NID g1261120  
VERSION N85495.1 GI:1261120  
KEYWORDS EST.  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 331)  
AUTHORS Liew,C.C.  
TITLE cDNAs from fetal heart (1996)  
JOURNAL Unpublished (1996)  
COMMENT On May 8, 1995 this sequence version replaced gi:801073.  
Contact: Liew CC  
Department of Laboratory Medicine and Pathobiology  
University of Toronto  
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5  
Tel: 416788758  
Fax: 4169785650



This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:03:27 ; Search time 921.96 Seconds  
(without alignments)  
63.772 Million cell updates/sec

Title: US-09-049-696-12

Perfect score: 235  
Sequence: 1 GACACCGACGAATTCGCCAG.....CAGTGAAGAATGCGGCGTC 235

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	27.7	878	1 T45884	Human colon specif
2	65	27.7	878	1 V16672	Polynucleotide seq
3	31.4	13.4	531	1 T30073	Thermostable Inorg
4	31	13.2	844	1 X40000	Prostate cancer as
5	30.8	13.1	579	1 O73872	Borrelia W antigen
6	30.8	13.1	528	1 O73897	Borrelia W antigen
7	30	12.8	6901	1 T78853	Human lectin-cho
8	29.8	12.7	3342	1 V52339	Streptococcus pneu
9	29.4	12.5	1127	1 X39991	Prostate cancer as
10	29.4	12.5	773	1 X39992	Prostate cancer as
11	29.2	12.4	1663	1 X39662	Renal cancer assoc
12	28.2	12.0	688	1 Q21768	Sequence encoding
13	28	11.9	1122	1 O41265	Modified protease
14	27.8	11.8	428	1 V04576	Flea serine protea
15	27.8	11.8	428	1 V04577	Flea serine protea
16	27.8	11.8	841	1 V04578	Flea serine protea
17	27.8	11.8	717	1 V04580	Flea serine protea
18	27.8	11.8	717	1 V04581	Flea serine protea
19	27.8	11.8	841	1 V04579	Flea serine protea
20	27.8	11.8	1800	1 V27582	Helicobacter pylor
21	27.8	11.8	110000	1 V21209..00	Methanococcus jann
22	27.6	11.7	15219	1 V18277	RSV vaccine 2B33F
23	27.6	11.7	15219	1 V18278	RSV vaccine 2B33F
24	27.6	11.7	15219	1 V18279	RSV vaccine 2B33F
25	27.6	11.7	15219	1 V18280	RSV vaccine 2B33F
26	27.6	11.7	15219	1 V18275	RSV vaccine 2B33F
27	27.6	11.7	15219	1 X35267	DNA encoding the L
28	27.6	11.7	15219	1 X35269	DNA encoding the L
29	27.6	11.7	15219	1 X35270	DNA encoding the L
30	27.6	11.7	15219	1 X35271	DNA encoding the L
31	27.6	11.7	15219	1 X35272	DNA encoding the L
32	27.6	11.7	2899	1 T88746	Streptococcus bovi
33	27.2	11.6	315	1 T20005	Human gene signatu
34	27.2	11.6	1485	1 X07184	Soybean S-adenosyl
35	27	11.5	3280	1 N82053	DNA encoding fibro
36	27	11.5	288	1 T23245	Human gene signatu
37	27	11.5	110000	1 V21209..06	Continuation (7 of
38	27	11.5	13740	1 V74365	Staphylococcus aur
39	26.8	11.4	1080	1 T13046	Cotton fibre-speci
40	26.8	11.4	1080	1 T130263	Cotton fibre-speci
41	26.8	11.4	1080	1 T62622	Cotton fibre-speci
42	26.8	11.4	1080	1 T70053	Cotton fibre-speci
43	26.8	11.4	1956	1 T67161	Plasmodium falcipa

44 26.8 11.4 110000 1 V21209..10  
45 26.8 11.4 10813 1 V74675

Continuation (11 o  
Staphylococcus aur

## ALIGNMENTS

RESULT	1	
T45884		
ID	T45884 standard; cDNA; 878 BP.	
AC	T45884;	
DT	13-MAR-1997 (first entry)	
DE	Human colon specific gene CSG5 cDNA partial clone.	
KW	Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;	
KW	gene therapy; ss.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	cds	2..691
FT		/*tag= a
PN	MO9639419-A1.	
PD	12-DEC-1996.	
PF	06-JUN-1995; U07289.	
PA	06-JUN-1995; MO-U07289.	
PR	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Yu G;	
DR	WPI: 97-043054/04.	
DR	P-PSDB: W06548.	
PT	Human colon specific genes and their expression products - detection	
PT	of which, in non-colon tissue samples, can be used as indication of	
PS	colon cancer metastasis	
PS	Claim 1: Fig 5; 60pp; English.	
CC	13 cDNA clones (T45880-92), most of them partial clones, correspond	
CC	to human colon specific genes, designated CSG1, CSG2, etc., that	
CC	are primarily expressed in tissues derived from the colon. CSG7	
CC	and CSG10 show reduced expression in colon cancer cells as compared	
CC	to that in normal cells; the remaining genes are overexpressed in	
CC	colon cancer. The partial cDNA sequences can be used to isolate	
CC	full-length clones and genomic clones including the complete gene.	
CC	CSG nucleic acids can be used to produce CSG polypeptides (see also	
CC	W06545-53) in transformed host cells, as probes to detect disorders	
CC	of the colon, partic. colon cancer and colon cancer metastasis, and	
CC	in gene therapy.	
SO	Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;	
Query Match		
Best local Similarity 100.0%; Pred. No. 9.3e-13;		
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	171 TGTCTACTCAGGATTTTACACGATGAGTGTAGTACAGTGAAGTGGC 230	
DB	1 TGTCTACTCAGGATTTTACACGATGAGTGTAGTACAGTGAAGTGGC 60	
QY	231 GGCTC 235	
DB	61 GGCTC 65	
RESULT 2		
ID	V16672 standard; cDNA; 878 BP.	
AC	V16672;	
DT	22-JUN-1998 (first entry)	
DE	Polynucleotide sequence of a colon-specific gene.	
KW	Colon-specific gene; probe; detection; expression; human;	
KW	diagnostic assay; colon cancer; antibody; screening; ss.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	cds	2..685
FT		/*tag= a
PN	US5733748-A.	
PD	31-MAR-1998.	
/*note= "no stop codon given"		

PF 06-JUN-1995; 469667.  
PR 06-JUN-1995; US-469667.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen C, Yu G;  
DR WPI: 98-229823/20.  
P-PSDB: W46879.  
PT Colon-specific nucleic acids - useful as probes for detecting colon  
PT cancer micrometastases  
PS Claim 15; Fig 5A-B; 51pp; English.  
CC V16666-81 represent polynucleotide sequences of partial or full-length  
CC cDNA clones of colon-specific genes. The polynucleotides can be used  
CC as probes to detect expression of the corresponding human genes, e.g. in  
CC diagnostic assays for detecting micrometastases of colon cancer.  
CC Recombinant cells containing the polynucleotides can be used to  
CC produce the polypeptides, in order that antibodies can be raised and  
CC used in further screening or diagnostics.  
SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 27.7%; Score 65; DB 1; Length 878;  
Best Local Similarity 100.0%; Pred. No. 9,3e-13;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 171 TGTCTACTCAAGTATTTCACAACTTATGACAGCAATGTAGATACGTAAAGTGGC 230  
DB 1 TGTCTACTCAAGTATTTCACAACTTATGACAGCAATGTAGATACGTAAAGTGGC 60  
OY 231 GGCTC 235  
DB 61 GGCTC 65

RESULT 3  
ID T30073 standard; DNA: 531 BP.  
AC T30073;  
DT 06-JAN-1997 (first entry)  
DE Thermolabile inorganic pyrophosphatase gene.  
KW Inorganic pyrophosphatase; pyrophosphate phosphohydrolase; Ppase;  
KW thermolabile enzyme; ds.  
OS Thermococcus litoralis strain NS-C (DSM 5473).  
PN W09612798-A1.  
PD 02-MAY-1996.  
PF 23-OCT-1995; U13662.  
PR 25-OCT-1994; US-329721.  
PA (NEWE) NEW ENGLAND BIOLABS INC.  
PI Lennox T, Sears LE, Slatko BE;  
DR WPI: 96-230599/23.  
P-PSDB: R88153.  
PT Thermolabile inorganic pyro:phosphatase isolated from Thermococcus  
PT litoralis - catalyses the hydrolysis of inorganic pyro:phosphate  
PT Claim 5; Page 57; 85pp; English.  
CC An isolated gene sequence (T30073) codes for the inorganic  
CC pyrophosphatase (Ppase) (R88153) of Thermococcus litoralis. It was  
CC obtd. by PCR amplification of T. litoralis genomic DNA using  
CC primers 1092 + 1095, 1093 + 1095 and 1094 + 1095 (see also T30069-  
CC 73). The gene can be incorporated into an expression vector to  
CC allow prodn. of highly purified Ppase in transformed host cells.  
CC The Ppase can be used to eliminate problems of pyrophosphorolysis  
CC e.g. in thermal cycling reactions.  
SQ Sequence 531 BP; 176 A; 110 C; 131 G; 114 T;

Query Match 13.4%; Score 31.4; DB 1; Length 531;  
Best Local Similarity 56.2%; Pred. No. 0.21;  
Matches 59; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 8 GCAATTCCTCCAGCCCTCTGCTAGTTTGAATATTCGCCAAGAGCCCTCCCAATTC 67  
DB 262 GCAAGACCAATATGAGCCTCTTCAGATGATAGACAGCGCGAAGAGCTACAGATATG 321  
OY 68 TCAGGGCCAGTGTCTACAGCCCTGATGATGATGATGATGATGATGATGATGATGAT 112  
DB 111 TATTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT

DB 322 GCAGTTCAGTGAAGATCCCTACTTATGACTGAGAGACATA 366

## RESULT 4

ID X40000 standard; DNA: 844 BP.  
AC X40000;  
DT 02-JUL-1999 (first entry)  
DE Prostate cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
OS Homo sapiens.  
PN W09904265-A2.  
PD 28-JAN-1999.  
PF 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.  
PR 10-OCT-1997; US-948705.  
PR 11-OCT-1997; GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, Ohtani Y, Old LJ,  
PI Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E,  
PI Tureci O;  
DR WPI: 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67; Page 629; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 844 BP; 251 A; 196 C; 234 G; 156 T;

Query Match 13.2%; Score 31; DB 1; Length 844;  
Best Local Similarity 53.8%; Pred. No. 0.34;  
Matches 64; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

OY 82 ACAGCCCTGATTAATCAAGTGAAGAAACAGTTCCTGTAAGTATGATGA 141  
DB 552 ACACCCATTAAGAAACAGGATGAGAAAGTAACTTTTGAACCTCTCTGGG 611  
OY 142 GCAGGTCTATCTCTAAGATGACGGTCTCTACTCAAGTATTTCACAACTATGA 200  
DB 612 GATGAATAATGGAGCTAGTATAAAGAGATGATTCAGATCCCTATCTATGATCA 670

## RESULT 5

ID 073872 standard; DNA: 579 BP.  
AC 073872;  
DT 25-MAY-1995 (first entry)  
DE Borrelia W antigen vaccine.  
KW OspC antigen; vaccine; Lyme disease; borreliosis; immunogen;  
KW serovar typing; restriction fragment length polymorphism;  
KW RFLP; Pichia pastoris; ss.  
OS Borrelia burgdorferi W.  
PN W09425596-A.  
PD 10-NOV-1994.

ET /tag= p









PT cats or dogs  
 PS Claim 1; Page 207; 318pp: English.  
 CC The present sequence encodes a novel flea serine protease. The  
 CC protease, its mimetopes, antibodies (Ab) and inhibitors of the  
 CC protein, as well as the DNA encoding the protein, may all be used in  
 CC therapeutic compositions to reduce flea infestation, especially in  
 CC immunoglobulin protease) and so reduce flea infestation, especially in  
 CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult  
 CC fleas which comprises anti-protease antibodies produced by a host animal  
 CC in response to administration of the protein. Therapeutic compositions  
 CC may further comprise a compound that reduces haematophagous ectoparasite  
 CC burden by a method other than by reducing flea immunoglobulin protease  
 CC activity. The novel flea DNA encoding the protein can also be used to  
 CC produce recombinant protein, and fragments of it are used as probes and  
 CC primers for identification and isolation of related sequences, also as  
 CC antisense, triplex-forming agents and ribozymes for inhibition of the  
 CC synthesis of the protein. Ab are also useful for screening expression  
 CC libraries, to purify the protein and to target cytotoxins to fleas.  
 SQ Sequence 428 Bp; 142 A; 78 C; 102 G; 105 T;

Query Match 11.8%; Score 27.8; DB 1; Length 428;  
 Best Local Similarity 55.8%; Pred. No. 3.1;  
 Matches 53; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 98 CAGTGAATGAAAAACAGTTACTTGAGACTGATGATGAGACAGGCTGATGCTA 157  
 DB 209 CAGAACTGAAAAAGAGTTCCAGAAAGCGCACTGTTAGTGCACAGATGGGGGCTA 268  
 QY 158 CTAGAGATGACGGTGTCTCTCAAGTATTTCACA 192  
 DB 269 CTATGTTGGGGCCCGCCAGCATCAACGCATCTATAAA 303

## RESULT 15

V04577/c  
 ID V04577 standard; DNA; 428 Bp.

AC V04577;

DT 02-JUL-1998 (first entry)

DE Flea serine protease DNA sequence SEQ ID NO:39.

KW Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;

OS Immunoglobulin protease; larvae; host animal; ss.

PN Siphonaptera.

WO9740058-A1.

PD 30-OCT-1997.

PF 24-APR-1997; U06121.

PR 04-APR-1997; US-042945.

PR 24-APR-1996; US-639075.

PR 15-NOV-1996; US-749699.

PA (HESK-) HESKA CORP.

PI Frank GR, Gaines PJ, Griewe RB, Hunter SW, Rushlow KE,

PI Silver G, Stiegler GL;

DR WPI: 98-076762/07.

PT New flea protease genes and proteins - used in vaccine compositions  
 for the prophylaxis and treatment of flea infestation, especially in  
 cats or dogs

PS Claim 1; Page 208; 318pp: English.

CC The present sequence is a DNA sequence from a novel flea serine protease.

CC The protease, its mimetopes, antibodies (Ab) and inhibitors of the

CC protein, as well as the DNA encoding the protein, may all be used in

CC therapeutic compositions to reduce flea protease activity (especially

CC immunoglobulin protease) and so reduce flea infestation, especially in

CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult

CC fleas which comprises anti-protease antibodies produced by a host animal

CC in response to administration of the protein. Therapeutic compositions

CC may further comprise a compound that reduces haematophagous ectoparasite

CC burden by a method other than by reducing flea immunoglobulin protease

CC activity. The novel flea DNA encoding the protein can also be used to

CC produce recombinant protein, and fragments of it are used as probes and

CC primers for identification and isolation of related sequences, also as

CC antisense, triplex-forming agents and ribozymes for inhibition of the

CC synthesis of the protein. Ab are also useful for screening expression

CC libraries, to purify the protein and to target cytotoxins to fleas.

SQ Sequence 428 Bp; 105 A; 102 C; 78 G; 142 T;

Query Match 11.8%; Score 27.8; DB 1; Length 428;

Best Local Similarity 55.8%; Pred. No. 3.1;

Matches 53; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 98 CAGTGAATGAAAAACAGTTACTTGAGACTGATGATGAGACAGGCTGATGCTA 157  
 DB 220 CAGAACTGAAAAAGAGTTCCAGAAAGCGCACTGTTAGTGCACAGATGGGGGCTA 161  
 QY 158 CTAGAGATGACGGTGTCTCTCAAGTATTTCACA 192  
 DB 160 CTATGTTGGGGCCCGCCAGCATCAACGCATCTATAAA 126

Search completed: August 6, 1999, 04:03:30  
 Job time: 1562 sec

This Page Blank (uspto)

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:32:58 ; Search time 3075.15 Seconds  
(without alignments)  
185.377 Million cell updates/sec

Title: US-09-049-696-7

Perfect score: 289  
Sequence: 1 GAAATATCCAACTGATGGAT.....AGCTTGAGAGTAAGGATTA 289

Scoring table:

Searched: 2546578 seqs, 986266752 residues

Database: EST:\*

1: em\_est1:\*\n2: em\_est2:\*\n3: em\_est3:\*\n4: em\_est4:\*\n5: em\_est5:\*\n6: em\_est6:\*\n7: em\_est7:\*\n8: em\_est8:\*\n9: em\_est9:\*\n10: em\_est10:\*\n11: em\_est11:\*\n12: em\_est12:\*\n13: em\_est13:\*\n14: em\_est14:\*\n15: em\_est15:\*\n16: em\_est16:\*\n17: em\_est17:\*\n18: em\_est18:\*\n19: em\_est19:\*\n20: gb\_est1:\*\n21: gb\_est2:\*\n22: gb\_est3:\*\n23: gb\_est4:\*\n24: gb\_est5:\*\n25: gb\_est6:\*\n26: gb\_est7:\*\n27: gb\_est8:\*\n28: gb\_est9:\*\n29: gb\_est10:\*\n30: gb\_est11:\*\n31: gb\_est12:\*\n32: gb\_est13:\*\n33: gb\_est14:\*\n34: gb\_est15:\*\n35: gb\_est16:\*\n36: gb\_est17:\*\n37: gb\_est18:\*\n38: gb\_est19:\*\n39: gb\_est20:\*\n40: gb\_est21:\*\n41: gb\_est22:\*\n42: gb\_est23:\*\n43: gb\_est24:\*\n44: gb\_est25:\*\n45: gb\_est26:\*\n46: gb\_est27:\*\n47: gb\_est28:\*\n48: gb\_est29:\*\n49: gb\_est30:\*\n50: gb\_est31:\*\n51: gb\_est32:\*\n52: em\_est20:\*\n53: em\_est21:\*

54: em\_est22:\*\n55: em\_est23:\*\n56: em\_est24:\*\n57: em\_est25:\*\n58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	77	26.6	468	37	AA726662	AA726662 vus3f05.r
2	70.8	24.5	337	50	AV047614	AV047614 AV047614
3	69.4	24.0	435	33	AA443258	AA443258 aa14d09.r
4	59.6	20.6	455	39	AA871197	AA871197 vq32c11.r
5	55.6	19.2	207	41	AF071903	AF071903 AF071903
6	38.2	13.2	430	47	A1493356	A1493356 tg70f11.x
7	36.4	12.6	587	44	AU044136	AU044136 AU044136
8	31.8	11.0	691	49	AU057081	AU057081 AU057081
9	31.6	10.9	221	32	AA367484	AA367484 EST78591
10	31.4	10.9	410	47	A1483251	A1483251 EST242728
11	31.2	10.8	520	43	A1187452	A1187452 EST240 Ma
12	31	10.7	267	20	T29364	T29364 EST78583 Hu
13	31	10.7	506	24	H91905	H91905 y881f08.r1
14	31	10.7	184	32	AA367525	AA367525 EST78604
15	31	10.7	473	46	A1437409	A1437409 fb33c09.x
16	30.8	10.7	518	40	AA931711	AA931711 o138e10.s
17	30.6	10.6	771	41	A1007148	A1007148 ua81f12.r
18	30.4	10.5	440	26	W45744	W45744 mc79e12.r1
19	30.4	10.5	209	32	AA367483	AA367483 EST78590
20	30.4	10.5	182	32	AA367519	AA367519 EST78604
21	30.4	10.5	422	50	A1702579	A1702579 we80b12.x
22	30.2	10.4	287	21	T79052	T79052 vq17b02.r1
23	30.2	10.4	371	25	N88412	N88412 K3341F Huma
24	30.2	10.4	529	37	AA692991	AA692991 vr60e09.s
25	30.2	10.4	266	48	A1616242	A1616242 vr60e09.x
26	30	10.4	527	30	AA198534	AA198534 mu21f03.r
27	29.8	10.3	427	27	AA041606	AA041606 m106h09.r
28	29.8	10.3	457	27	AA048018	AA048018 m123c05.r
29	29.8	10.3	436	27	W83004	W83004 m122d04.r1
30	29.8	10.3	256	33	AA403668	AA403668 va31a08.r
31	29.8	10.3	328	33	AA444680	AA444680 ve73d04.r
32	29.8	10.3	561	34	AA467006	AA467006 ve71c03.r
33	29.6	10.2	209	26	W41881	W41881 mc64g06.r1
34	29.6	10.2	351	27	AA014600	AA014600 mg94a01.r
35	29.6	10.2	389	27	AA030443	AA030443 m12e11.r
36	29.6	10.2	292	27	AA030469	AA030469 m141h03.r
37	29.6	10.2	546	27	W83098	W83098 m101a03.r1
38	29.6	10.2	338	29	AA139876	AA139876 m72d11.r
39	29.6	10.2	377	33	AA387117	AA387117 vc32e01.r
40	29.6	10.2	362	33	AA387129	AA387129 vc32a01.r
41	29.6	10.2	426	33	AA387152	AA387152 vc22g01.r
42	29.6	10.2	566	38	AA796769	AA796769 vp27d04.r
43	29.6	10.2	910	43	A1227295	A1227295 u108a12.y
44	29.6	10.2	526	48	A1587742	A1587742 mp33b08.y
45	29.6	10.2	685	51	A1737001	A1737001 sb36d10.y

#### ALIGNMENTS

RESULT 1  
AA726662 468 bp mRNA EST 02-JAN-1998  
LOCUS vus3f05.r1 Striatogene mouse skin (#937313) Mus musculus CDNA clone  
DEFINITION IMAGE:1209729 5' similar to SW:EDIC\_BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ; , mRNA sequence.  
ACCESSION AA726662  
NID 92744369

VERSION AA726662.1 GI:2744369  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 468)  
 AUTHORS Maria, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On May 5, 1995 this sequence version replaced gi:798073.

CONTACT: Maria M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:646073  
 Seq primer: -28m13 rev1 ET from Amersham  
 High quality sequence stop: 443.

FEATURES  
 source  
 Location/Qualifiers  
 1..468  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone\_image="1209729"  
 /clone\_lib="Stratagene mouse skin (#937313)"  
 /sex="females"  
 /tissue\_type="whole skin"  
 /dev\_stage="11 weeks old"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: skin; Vector: pBluescript SK-; Site: 1:  
 EcoRI, Site: 2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. Whole skin from 11 week old C57BL/6 female mice.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' CTCGAGTGTGTTTGTGTTTGTGTTT 3' adaptor  
 sequence: 5' CTCGAGTGTGTTTGTGTTTGTGTTT 3'."

BASE COUNT 121 a 107 c 120 g 120 t  
 ORIGIN

Query Match 26.6%; Score 77; DB 37; Length 468;  
 Best Local Similarity 55.2%; Pred. No. 8.2e-15;  
 Matches 149; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

OY 10 AACGATGATCTGAATGTCGCTGCTGACGATGGGGAACAACACTATTAATGGGTG 69  
 |||||  
 DB 30 AGCTGACGGCTGCTGCTGATTAATGATGACGACATGATGAACATGATGCAACTG 89  
 |||||  
 OY 70 CTTTAACGAGGTCAAAAGTNGTCATCATCCACACAGTCGTTGGGGCCCTGTC 129  
 |||||  
 DB 90 CCTGCTACCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 149  
 |||||  
 OY 130 AGCTCAAGACTAGAGAGAGTGTCCAAATGACAGAGGTTTACAGACATATGCTTCA 189  
 |||||  
 DB 150 AGCCAGAAAGTGGGGATATATACGCTTACAGAGAGGCTTAAGTTCTTCATTCAGA 209  
 |||||  
 OY 190 TCAAGTTCAGAACATGCGCTCATGATGCTTTGGGGCCCTTTCATCAGGAATGAGC 249  
 |||||  
 DB 210 TAAATTTACTTCTATGGAATGACTGAGCTTGTGCTGCAATCTCTTGAACAGAGA 269  
 |||||  
 OY 250 TGTCTCAGCGCTCATCCAGCTTGAGAG 279  
 |||||  
 DB 270 CATTTCCAGCAAGCTTACAGGTGAGAG 299

RESULT 2  
 LOCUS AV047614/c  
 DEFINITION AV047614 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone  
 170069H09, mRNA sequence.  
 ACCESSION AV047614  
 NID 94867279  
 VERSION AV047614.2 GI:4867279  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 337)  
 AUTHORS Akahira, S., Shibata, K., Ozawa, Y., Konno, H., Ito, M., Aizawa, K.,  
 Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,  
 Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H.,  
 Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,  
 Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,  
 Tomimaga, N., Watanabe, S., Yagame, M., Yamamata, T., Yokota, T.,  
 Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
 TITLE Riken Mouse ESTs  
 JOURNAL Unpublished (1999)  
 COMMENT On Feb 17, 1998 this sequence version replaced gi:2889739.

CONTACT: Chie Owa  
 Genome Science Laboratory  
 RIKEN  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-9145  
 Fax: 81-298-36-9098  
 Email: genome-res@rtc.riken.go.jp  
 Thermostabilization and thermocycling of thermolabile enzymes by  
 trehalose and its application for the synthesis of full length cDNA  
 (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
 Transcriptional sequencing: A method for DNA sequencing using RNA  
 polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
 further details.

FEATURES  
 source  
 Location/Qualifiers  
 1..337  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_image="170069H09"  
 /clone\_lib="Mus musculus adult C57BL/6J testis"  
 /sex="male"  
 /tissue\_type="adult"  
 /dev\_stage="testis"

BASE COUNT 103 a 67 c 72 g 95 t  
 ORIGIN

Query Match 24.5%; Score 70.8; DB 50; Length 337;  
 Best Local Similarity 69.2%; Pred. No. 7.4e-13;  
 Matches 110; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

OY 11 ACTGATGATCTGAATGTCGCTGCTGACGATGGGGAACAACACTATTAATGGGTG 70  
 |||||  
 DB 159 ACTTCGGTTCGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 101  
 |||||  
 OY 71 TTTAAGAGGTCAAAAGTNGTCATCATCCACACAGTCGTTGGGGCCCTTCA 130  
 |||||  
 DB 100 TTTAAGAGGTCTTCGCGCGCGCTGATCATCATCATCATCATCATCATCATCATCAT 141  
 |||||  
 OY 131 GCTCAAGAGTAGAGAGGTGTCCAAATGACAGAGGT 169  
 |||||  
 DB 40 GCCCTAAACTGAAATCTGTCGAAATGAAAGTAGT 2

RESULT 3

	AA443258				
LOCUS	AA443258	435 bp	mRNA	EST	03-JUN-1997
DEFINITION	aa14d09.r1 Soares.NHMPU.S1 Homo sapiens cDNA clone IMAGE:813223 5'				
	similar to T7:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL				
	; , mRNA sequence.				
ACCESSION	AA443258				
NID	G2155933				
VERSION	AA443258.1	GI:2155933			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 435)				
AUTHORS	Hallier,T., Allen,M., Bowles,D., Dubucque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wille,T., Waterston,R. and Wilson,R. WashU-Merck EST Project 1997 Unpublished (1997)				
TITLE	JOURNAL				
COMMENT					

**CONTACT:** Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel.: 314 266 1800  
 Tel.: 314 266 1800  
 Fax: 314 266 1810  
 Email: [estewatson.wustl.edu](mailto:estewatson.wustl.edu)  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
 Seq primer: -26m3 rev2 ET from Amersham  
 High quality sequence stop: 412.  
 Location/Qualifiers

```

FEATURES
source
    Location/Qualifiers
        1..435
            /organism="Homo sapiens"
            /db_xref="GDB:6044143"
            /db_xref="taxon:9606"
            /clone IMAGE:813233"
            /clone_1lb="Scares_NHMPu_S1"
            /tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
            /lab_host="DH10B"
            /note="Organ: mixed (see below): Vector: pTT73D-Pac
(Parmacia) with a modified polylinker; Site_1: Not I.
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHb, pregnant uterus
NBHbU, and fetal heart NBHH19W) were mixed, and ss clones
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484468-489479."
BASE COUNT
131 a      86 c      86 g      132 t
ORIGIN

```

Query Match	24.08;	Score	69.4;	DB	33;	Length	435;
Best Local Similarity	63.58;	Pred.No.	2.3e-12;				
Matches 106; Conservative	0;	Mismatches	61;	Indels	0;	Gaps	0

Qy	119	GGGCGCTTCAGCTCAAGAAGCTAGAGAGCTGTCACAAAATGACGAGAGTTTACACACA	178
Db	1	GGTTCATCTGCAGCCCAATCTGGAGGAATTATTCAGCTCTACAGGAGGTTTAAAGTTC	60
Qy	179	TATGCTTCAGATCAAGTTTCACAAAGAAGCCTCAATGATGTTTGGGCGCCTTATCA	238
Db	61	TTTGTTCCAGATATATATCAAACTCCAAATAGCTATGATGTTTACATAGAAATTCTCT	120
Qy	239	GGAAATGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAAGATAAGG	285
Db	121	GGAACTGGAGACATTTTCCACACCAACATATTCAGCTTGAAGATAAGG	167

RESULT 4  
LOCUS AA871197  
DEFINITION AAC871197 455 bp mRNA EST 16-MAR-1998  
IMAGE:109556 5' similar to SM:EDLC.BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.  
ACCESSION AA871197  
NID 92966642  
VERSION AA871197.1 GI:2966642  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 455)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubug  
geisel,S., Kucab,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The Mashu-HMT Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1395394.

**TITLE**  
 The WashU-HMI Mouse EST Project  
**JOURNAL**  
 Unpublished (1996)  
**COMMENT**  
 On Sep 12, 1996 this sequence version replaced gi:1395394.  
  
**Contact:** Maria M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LIND ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:602188  
 Seq primer: -28mJ rev2 ET from Amersham  
 High quality sequence stop: 414.  
 Location/Qualifiers

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="tacon:10090"
/clone_image="IMAGE:1095956"
/clone_id="Birstead bowel MPURB9"
/tissue_type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCAGCATCTGTAAGGAGGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[TAATCGATCCTTGG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Source irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob
Birstead."

```

Query Match	20.6%	Score 59.6	DB 39	Length 455
Best Local Similarity	78.9%	Pred. No. 3.5e-09		
Matches	71	Conservative	0	Mismatches 19; Indels 0; Gaps 0
QY 199	GAACATGGCCATCATGATGCTTGGGGCCCTTCAATCAGAAATGAGACTGTCCTCA	258		
Db 4	GAACAAAGGCTGTGTGATGCTTGGCAGCCACTCTCTCTCCAGAAATGCGCGATGCTCA	63		
QY 259	GGCGTCATCCAGCTTGAGAGATTAAGGATT	288		
Db 64	GCATCTCATCCAGCTGAGAGAGCAGGAGGT	93		

RESULT	5				
LOCUS	AF071903/C				
DEFINITION	AF071903	207 bp	mRNA	EST	30-JUN-1998
ACCESSION	AF071903	Fetal sorted endothelial cell line CPA47	Bos taurus CDNA clone 38, mRNA sequence.		
NID	93265169				
VERSION	AF071903.1	GI:3265169			
KEYWORDS	EST.				
SOURCE	Bos taurus.				
ORGANISM	Bos taurus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae Bovinae; Bos.				
AUTHORS	1 (bases 1 to 207)				
TITLE	Oberst,T.B. and Allegro,M.C.				
JOURNAL	Pigpen-binding mRNA Unpublished (1998)				
COMMENT	On May 18, 1995 this sequence version replaced gi:811149.				

**FEATURES**

**Contact:** Mark C. Alliegro

**Anatomy**

Louisiana State University Medical Center  
1901 Perdido Street, New Orleans, LA 70112, USA  
Email: malle@lsunc.edu.

**Location/Qualifiers**

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/map="1"
/clone="38"
/clone_lib="Fetal aorta endothelial cell line CPA47"
/tissue_type="aorta"
/cell_type="endothelium"
/cell_line="CPA47"
/dev_stage="fetus"
/note="p1gpen (GenBank Accession Number U26024) -binding
mRNA"

```

	Query Match	19.2%	Score 55.6;	DB 41;	Length 207;
	Best Local Similarity	60.0%	Pred. No. 5.3e-08;		
	Matches 114;	Conservative 0;	Mismatches 69;	Indels 7;	Gaps 1.
QY	99 TCATCCACACAGTCGCTTTGGGGCCCTGCAAGCTCAAGAACTAGAGAGAGCTGTCCAAA	158			
Db	203 TCTAGCCCGCGGAGAGCTGTGGAGCCCTGCTGCGCAAGAACTGGAGACATTTGCAATA	144			
QY	159 TGACAGAGAGTTACGACACATATGCTTAGATCAAGTTCAACAAATAGGCTCATTTGATG	218			
Db	143 TGACAGAGAGATATCTTTTGTGCAATTAAG-----AATATAGGCGCTTACTTAAGC	91			
QY	219 CTTTTCGGGCCCCCTTTCATCAGGAATGAGAGCTGTCTCAGACGGCTCCATCAGCTTGAGA	278			
Db	90 CTTTCAGTGAATTTTCATCTAGAAAGTGGAGAGATACCTCAGACAGCTATTACGTGGAAA	31			
QY	279 GTTAGGGATT 288				
Db	30 GCAAAAGCTTT 21				

RESULT	6
A1493356/c	
LOCUS	A1493356
DEFINITION	430 bp mRNA EST
	Ig70f011.x1 Soares.NHnpnu.SI Homo sapiens CDNA clone IMAGE:2114157
	3 similar to SW:ECIC_BOVIN P54281 EPITHELIAL CHLORIDE CHANNEL
PROTEIN :	/ mRNA sequence.
ACCESION	A1493356
NID	G4394359
VERSION	A1493356.1
KEYWORDS:	GT:G4394359
EST	*

**SOURCE ORGANISM** human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 430)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**AUTHORS** National Cancer Institute, Cancer Genome Anatomy Project (CGAP)  
**TITLE** Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** On May 18, 1998 this sequence version replaced gi:3137059.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the IMACS Consortium (info@lmage.llnl.gov) for further information.  
Insert Length: 396 Std Error: 0.00  
Seq Primer: -40UP from Gibco  
High quality sequence stop: 367.  
Accession: F04115.1

FEATURES

source

Location/Qualifiers

1. 430

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="214157"

/clone\_lib="Sources\_vhMMPu\_S1"

/tissue\_type="Pooled human melanocyte, fetal heart, and pregnant uterus"

/lab\_host="DH10B"

/note="Organ: mixed (see below); Vector: pT7N3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBH9, pregnant uterus 2NBH9U, and fetal heart 2NBH19M) were mixed, and ss circles were made *in vitro*. Following NBP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.E. clones 260332-265223, 340488-345479, and 484488-489479."

136 a 84 c 130 t

BASE COUNT

ORIGIN

Query Match	Best Local Similarity	Score	DB	Length
Matches 55; Conservative	66.38;	38.2;	47;	430;
	Pred. No. 0.03;			
	Mismatches 28; Indels 0; Gaps 0;			
QY 203	AATGGCCATGATGATGCTTTGGGGCCCTTCACACAGAAATGAGCTGCTCTACGCC	262		
Db 430	AATAGCATGATTTGATGCTTTCAGTAGAATTTCTTCGAACTGGAGACATTTCCAGCAA	371		
QY 263	TCCATCCAGCTTGAGATGAGG	285		
Db 370	CATATTCAGCTTGAAGATCAGG	348		

RESULT	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
7	AU0044136/c		AU0044136	587 bp	RNA	EST			09-DEC-1998	
			AU0044136	Mouse sixteen-cell-embryo	CDNA	Mus musculus	CDNA	clone		
			J0915E04	3'	RNA sequence.					
			AU0044136							
			93979611							
			AU0044136.1	GI:3979611						
			EST.							
			house mouse.							
			Mus musculus							
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;							
			Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.							
			1 (bases 1 to 587)							
			KO,M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,							
			DePalma,G.E., Liang,Y., Kargul,G.J., Sharara,R., Lim,M.K. and							
			Dot,H.							



```

/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="S2113_1A"
/clone_1lb="Oryza sativa mature leaf Nipponbare"
/tissue_type="mature leaf"

```

```

/organism="Homo sapiens"
/db_xref="ATCC (inhost):172265"
/db_xref="taxon:9606"
/clone_id="Paratid gland tumor"
/dev stage="adult"
note="Organ: parotid gland; Vector: pBluescript SK-";

```



Db	300	AAACATTTACTATATAAATATTAAAGCTTAAACATACATACATATAAAGATTAAAGATTGACATCTCTGACACC	241
Oy	133	TCAGACTTACAGAGGAGGAGCTGTCCAAATATGACAGAGGTTTACAGACATATCTTCAGATCA	192
Db	240	TTACGATATATATATATAGCCATTCATAAATTAAGTTCAGATTAATAATATATCATCTTAAAAATA	181
Oy	193	A	193
Db	180	A	180

  

LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	
T29364	EST/7833 Human Salivary gland Homo sapiens CDNA 5' end similar to	T29364	9511462	T29364.1	GI:611462	EST.	Human.	Adams M.D., Kerlavage A.R., Fleischmann R.D., Fulton R.A., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Chui, M.-W., Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S.M., Glodex, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkley, P.S., Kelley, J.M., Kline, K.M., Kelley, J.C., Liu, L.-T., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pellegrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.-J., Dimke, D., Feng, P., Ferlie, A., Fischer, C., Hastings, G.A., He, W.-W., Hu, J.-S., Greene, J.M., Gruber, J.C., Hudson, P., Kim, A., Kozak, D.L., Kunsch, C., Ji, H., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Wei, Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haezel, W.A., Fields, C., Fraser, C.M., and Venter, J.C.	Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million Basepairs of cDNA Sequence	Nature 377, 3-174 (1995)
JOURNAL MEDLINE	96026280	Other_ESTS: THC24118	Contact: Venter, JC	The Institute for Genomic Research	932 Clapper Rd, Gathersburg, MD 20878	Tel: 3018699056	Fax: 3018699423	Email: tdbinfo@tdb.tigr.org	For clone availability, additional sequence and expression information related to this EST, please contact the TIGR Database (tdbinfo@tdb.tigr.org)	Seq primer: M13 Reverse.
FEATURES	Source	Location/Qualifiers	1..267	/Organism="Homo sapiens"	/db_xref="ATCC (inhost):106082"	/db_xref="taxon:9606"	/clone_lib="Human Salivary gland"	/note="Organ: salivary gland"		
BASE COUNT	69 a	70 c	69 g	53 t	6 others					
ORIGIN										

  

Query Match	Best Local Similarity	Matches
10.7%	Score 31; DB 20; Length 267;	52.3%; Pred No. 5.4;
67;	Conservative	0; Mismatches 61; Indels 0; Gaps 0

Oy	159	TGCAGGAGGTTCATACGACATATCGCTTACGACAGTCAGACAACTGGCCCATTTGATG	218
Dd	29	TGTCAATGGCCCCCTGTGGCCTTCAGCTACGCTCAGAGATTTAATGACATNTCACGCCAG	88
Oy	219	CTTTTGAGGCCCTTTCATCAGGAATGAGAGCTGTCTCTACGCGCTCATCCAGCTTGAGA	278
Dd	89	AAGATGTTCCCTCCTGTAATATCAGATGAGAGAGACTCTGACGAGTTTCATGATGAGAGAC	148
Oy	279	GTAAGGGA	266
Dd	149	GTCAGGGA	156

  

RESULT	13
LOCUS	H91905
DEFINITION	yaeli08.r1 Soares retina N2b4HR Homo sapiens CDNA clone
ACCESSION	IMAGE:221223 5', mRNA sequence.
NID	H91905
VERSION	g1087483
KEYWORDS	H91905.1 GI:1087483
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 506) Hillier,L., Clark,N., Duboue,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le-M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasxis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R. The WashU-Merck EST Project Unpublished (1995) On Sep 21, 1992 this sequence version replaced gi:276438.

  

TITLE	Contact: Wilson RK
JOURNAL	Washington University School of Medicine
COMMENT	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu High quality sequence stops: 299 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert length: 1517 Std Error: 0.00 Seq primer: ML3RP1 High quality sequence stop: 299. Location/Dualifiers 1. 506 /organism="Homo sapiens" /db_xref="GDB:3849216" /db_xref="taxon:9606" /map="8" /clone="IMAGE:221223" /clone_lib="Soares retina N2b4HR" /sex="male" /tissue_type="retina" /dev_stage="55 year old" /lab_host="DH10B (ampicillin resistant)" /note="Organ: eye; Vector: pRT73D (Pharmacia) with a modified polylinker; Site.1: Not I - Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCACATCGTAGAGTGGAGCGCGCCCTTTTCTTTTTTTT 3'] , reverse-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento

BASE COUNT 139 a 97 c 120 g 137 t 13 others  
 ORIGIN

Query Match 10.7% Score 31; DB 24; Length 506;  
 Best Local Similarity 51.0%; Pred. No. 6.8;  
 Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 123 CCTCTCAGCTCAGACATAGAGAGCTGTCCAAAATGACAGAGGTTTACAGACATATG 182  
 |||||  
 DB 167 CCTCTCAGCTCAGACATAGAGAGGTTTACAGACATATG 226  
 |||||  
 QY 183 CTTTCATCAGTTCAGACATAGGCTCTATTATGCTTTGGGCCCTTTCATCAGAA 242  
 |||||  
 DB 227 CTTGACAGCAATTTTCATGCGCAGTTCAGAGTATTTCATAGGGCTTTACTGTATCA 286  
 |||||  
 QY 243 ATGACGCTGTCTCTCAGCGCTC 265  
 |||||  
 DB 287 CTTTGATTTCCACAGCTTTTC 309

RESULT 14  
 AA367525 184 bp mRNA EST 21-APR-1997  
 LOCUS EST18611 Paratid gland tumor Homo sapiens cDNA 5' end similar to  
 DEFINITION similar to proline-rich protein Haeti, subfamily 1, mRNA sequence.  
 ACCESSION AA367525  
 NID 92019842  
 VERSION AA367525.1 GI:2019842  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 184)  
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulmer, R.A.,  
 Bult, C.J., Lee, N.H., Kirkness, E.F., Welsch, K.G., Gocayne, J.D.,  
 White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-  
 Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, E., Fine, L.D.,  
 Fitzgerald, L.M., Fitzhugh, W.M., Friedman, L., Geoghegan, N.S.,  
 Glodet, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,  
 Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
 Moreno-Palanges, R.F., McDonald, L.A., Nguyen, D.T., Peligrino, S.M.,  
 Phillips, C.A., Ryder, S.E., Scott, J.L., Saudak, D.M., Shirley, R.,  
 Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,  
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
 Dunke, D., Feng, D.-F., Fertle, A., Fischer, C., Hastings, G.A.,  
 He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,  
 Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Melsner, P.S., Olsen, H.,  
 Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M.,  
 Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,  
 Fraser, C.M., and Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns  
 JOURNAL based upon 83 million nucleotide of cDNA sequence  
 MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)  
 COMMENT 96026280  
 On Sep 12, 1996 this sequence version replaced gi:1297389.  
 Other ESTs: THC178414  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tcdb/ngi/ngi.html>)  
 Seq primer: M13 Reverse.

FEATURES  
 SOURCE Location/Qualifiers  
 1..184  
 /organism="Homo sapiens"

/db\_xref="ATCC (inhost):172284"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Paratid gland tumor"  
 /dev\_stage="adult"  
 /note="Organ: parotid gland; Vector: pBluescript SK-;  
 Site-1: EcoRI; Site-2: XhoI"  
 BASE COUNT 46 a 46 c 46 g 45 t 1 others  
 ORIGIN

Query Match 10.7% Score 31; DB 32; Length 184;  
 Best Local Similarity 52.3%; Pred. No. 4.7;  
 Matches 67; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 159 TGACAGAGGTTTACAGACATATGCTCAGATCAGTTCAGACATAGGCTCATGATG 218  
 |||||  
 DB 24 TGTCAATGGCCCTGTGCTCCTCAGCTCAGATTTAAATAGATGTGACCCAGG 83  
 |||||  
 QY 219 CTTTGGGGCCCTTTCATCAGAAATGAGCTGTCTCAGCGCTCCATCAGCTTGA 278  
 |||||  
 DB 84 AAGATGTTCCCTGTATATCAGATGAGAGAGACTGACAGCTCATATAGATGAGAGC 143  
 |||||  
 QY 279 GTAAGGA 286  
 |||||  
 DB 144 GTCAAGGA 151

RESULT 15  
 A1437409 473 bp mRNA EST 09-MAR-1999  
 LOCUS fd33c09.x1 zebrafish washu mping EST Danio rerio cDNA 3' similar to  
 DEFINITION SW:CI82.WAIZE Q10717 CYSTEINE PROTEINASE 2 PRECURSOR ; mRNA  
 sequence.  
 ACCESSION A1437409  
 NID 94285658  
 VERSION A1437409.1 GI:4285658  
 KEYWORDS EST.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
 Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
 Cyprinidae; Cyprinidae; Rasbora; Danio.  
 1 (bases 1 to 473)  
 Clark, M., Johnson, S.L., Lebach, H., Lee, R., Li, F., Marra, M.,  
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schur, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R., and Wilson, R.

TITLE Washu zebrafish EST Project 1998  
 JOURNAL Unpublished (1998)  
 COMMENT On May 18, 1998 this sequence version replaced gi:3138307.

Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: szbrafish@wustl.wustl.edu  
 CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone distribution: Genome Systems, St. Louis,  
 Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:  
[info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama  
 (web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and  
 RessourceZentrumPrimarDatenbank, Berlin, Germany (web address:  
[www.rzpd.de](http://www.rzpd.de))  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: T7 ER from Amersham  
 High quality sequence stop: 360.  
 Location/Qualifiers  
 1..473  
 /organism="Danio rerio"



**This Page Blank (uspto)**

Gencore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:02:57 ; Search time 921.96 Seconds  
(without alignments)  
68.657 Million cell updates/sec

Title: US-09-049-696-8

Perfect score: 253

Sequence: 1 AACCAAGTGTGTCATCATC.....GGCACATGATGTGACAG 253

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32.6	12.9	678	1	EST clone BK296. N
2	29	11.5	933	1	Sequence encoding
3	29	11.5	4821	1	T38305
4	28.6	11.3	414	1	T89065
5	28.6	11.3	7812	1	X24753
6	28.2	11.1	753	1	005733
7	28	11.1	371	1	T69307
8	28	11.1	3531	1	T80591
9	27.8	11.0	302	1	059984
10	27.8	11.0	110000	1	V21209_03
11	27.8	11.0	110000	1	V21209_04
12	27.6	10.9	3872	1	T10156
13	27.6	10.9	3872	1	X07394
14	27.6	10.8	8726	1	X13032
15	27	10.7	2457	1	T30736
16	27	10.7	2551	1	T30738
17	27	10.7	445	1	T30733
18	27	10.7	445	1	T30735
19	27	10.7	1633	1	T74810
20	26.8	10.6	8133	1	T29248
21	26.8	10.6	8133	1	V30560
22	26.8	10.6	5360	1	X13155
23	26.8	10.6	644	1	X20070
24	26.8	10.6	548	1	X20048
25	26.6	10.5	2305	1	039149
26	26.6	10.5	2311	1	039150
27	26.6	10.5	578	1	062811
28	26.6	10.5	5086	1	V68832
29	26.6	10.5	4232	1	X13008
30	26.4	10.4	3073	1	015280
31	26.4	10.4	9071	1	022491
32	26.4	10.4	8387	1	022489
33	26.4	10.4	9071	1	023370
34	26.4	10.4	2706	1	092972
35	26.4	10.4	110000	1	V30458_1
36	26.4	10.4	5549	1	V74499
37	26.4	10.4	477	1	V68416
38	26.4	10.4	110000	1	V30459_1
39	26.2	10.4	1581	1	N82012
40	26.2	10.4	455	1	V56686
41	26.2	10.4	487	1	V56684
42	26.2	10.4	478	1	V89283
43	26.2	10.4	1941	1	X24916

44 26.2 10.4 2264 1 X08433 Catalase gene of h  
c 45. 26 10.3 80073 1 T58840\_5 Continuation (6 of

## ALIGNMENTS

```
RESULT 1
V86928 V86928 standard; CDNA: 678 BP.
AC 27-APR-1999 (first entry)
DE EST clone BK296.
KW expressed sequence tag; secreted protein; hematopoiesis regulator;
KW tissue growth; actinin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN M09845435-X2.
PD 15-OCT-1998.
PF 10-APR-1998; U06954.
PR 10-APR-1997; US-835913.
PA (GEM) GENETICS INST INC.
PI Agostino MJ, Jacobs R, Lavallie ER, McCoy JM, Merberg D,
PI Racine LA, Spaulding V, Treacy M;
DR WPI: 99-070076/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1: Page 394-395: 633pp: English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC actinin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC therapy. The EST sequences are also stated to be useful for gene
CC sequence 678 BP: 120 A: 172 C: 198 G: 188 T:
SQ
Query Match 12.9% Score 32.6; DB 1; Length 678;
Best Local Similarity 48.2%; Pred. No. 0.13;
Matches: 92; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 63 GAGCTGTCCAAATGACAGAGGTTTACAGATATGCTTCAGATCAAGTTCAGAACAT 122
QY 11 ||||| 1 ||||| ||||| ||||| |||||
DB 263 GATTGTTCATTTGGGTGAGAGAGTTCAGTATATGTTGGATTTTTCAGGGTGA 322
QY 123 GGGCTCATTTGATCTTTGGGGCCCTTTTCATCAGAAATGAGACTCTCTCAGCCCTCC 182
QY 11 ||||| 1 ||||| ||||| ||||| |||||
DB 323 GGGATCGAGACTCTGGAGGGGACCCCTCTTCAGGACAGCGCCCTGCCCCAGTGCCTCC 382
QY 183 ATTCACCTTGAGAGTAAAGGATTACCCCTCCAGAACAGCAGTGAATGGCAGAGG 242
DB 383 ACTACCTTCCTGGGGGAGAACCTCAAGTCAAGCAGGCTGGGGACAGATTCCCCCA 442
QY 243 ATCGTGACAG 253
DB 443 AGCTGAGACAG 453
RESULT 2
QY 004495 004495 standard; DNA: 933 BP.
AC 004495:
DT 09-OCT-1990 (first entry)
```

DE Sequence encoding human bone osteonectin(ON).  
KW Osteoporosis; Rheumatoid arthritis; Paget's disease;  
KM atherosclerosis; periodontal; human bone matrix; proteoglycan.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 22..930  
FT /tag- a  
PN US7432044-A.  
PD 17-APR-1990.  
PF 3-NOV-1989; 432044.  
PR 3-NOV-1989; US-432044.  
PA (USSH) Nat Inst of Health.  
PI Termine J;  
DR WPI: 90-178641/23.  
P-PSDB: R05164.  
PT Human bone matrix DNA and proteins -  
PT used in detection, diagnosis and treatment involving skeletal  
PT and/or connective tissue disease states.  
PS Disclosure: P: English.  
CC Probes and Abs raised to the proteins can be used to determine  
CC their levels useful in diagnosis of associated connective tissue  
CC diseases states such as osteoporosis, osteo/rheumatoid arthritis,  
CC Paget's disease, atherosclerosis and periodontal disease.  
CC Proteins may also be used to induce or block biological function.  
SQ Sequence 933 BP; 226 A; 266 C; 265 G; 176 T;

Query Match 11.5%; Score 29; DB 1; Length 933;  
Best Local Similarity 45.7%; Pred. No. 2.4;  
Matches 101; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

OY 28 TCGCTTGGGGCCCTGTCAGCTCAGACTGAGAGCTGCTCCAAATGACAGAGGTT 87  
DB 53 TGGCCGGAGAGGCTTGGAGCCCTCCAGACAGAGCCCTGCTGATGACAGAGGTTG 112  
OY 88 TACGACATATGCTTCATGATCAATGACATGAGCCCTCATTTGCTTTGGGGCCC 147  
DB 113 TGGAGAACTGTCGACAGAGTGACTGATGCTGGAGAGCTATCTGTCAGAGTTG 172  
OY 148 TTTCATCAGAGAAATGAGCTGTCTCTCAGCCCTCCATCCAGCTTGAGAGTAAAGGATTAA 207  
DB 173 AAGTAGGAAATTTGATGATGTGTGTCAGAGAAACCCAGAGAGAGGTGTGGCGAAATTC 232  
OY 208 CCTCCAGACAGCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 248  
DB 233 CCTGCCAGAACCAACCACTCCAAACAGGAGAGGTGTGCGAG 273

## RESULT 3

T38305  
ID T38305 standard; DNA; 4821 BP.  
AC T38305.  
DT 03-JAN-1997 (first entry)  
DE Tospovirus M-RNA sequence.  
KM Tospovirus; disease resistance; crop protection; transgenic plant;  
KM tomato chlorotic spot virus; groundnut ringspot virus;  
KM Impatiens necrotic spot virus; viral movement protein; ds.  
OS Tomato spotted wilt virus.  
PN WO9629420-A1.  
PD 26-SEP-1996.  
PF 22-MAR-1996; E01271.  
PR 23-MAR-1995; GB-005907.  
PA (SANO ) SANDOZ LTD.  
PA (SANO ) SANDOZ PATENT GMBH.  
PA (SANO ) SANDOZ-ERFINDUNGEN VERW GMBH.  
PI De Haan PT;  
DR WPI: 96-443194/44.  
PT Recombinant nucleic acid contg. modified tospovirus sequence - used  
PT to prepare plants resistant or tolerant to tospovirus  
PS Claim 1; Page 14-16; 38pp; English.  
CC A DNA sequence (T38305) is complementary to one which encodes  
CC the tomato spotted wilt virus M-RNA sequence which itself encodes  
CC the putative viral movement protein. A recombinant nucleic acid

CC contains a transcriptional regulatory region and, under its  
CC control, the tospovirus sequence, pref. modified by removal or  
CC inactivation of the start codon, and insertion or deletion of  
CC nucleotides 3' of the start codon to alter the reading frame of the  
CC encoded RNA. A plant having such a construct stably integrated and  
CC expressed in its genome is resistant to, or tolerant of,  
CC tospoviruses.  
SQ Sequence 4821 BP; 1569 A; 850 C; 870 G; 1532 T;

Query Match 11.5%; Score 29; DB 1; Length 4821;  
Best Local Similarity 63.8%; Pred. No. 4.7;  
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

OY 97 ATGCTTCAGATCAAGTTCAGAACATGAGCTTCATGATGCTTTGGGGCCCTTCATCAG 156  
DB 476 AGTTTATATCACTTGAATAAACAATGCTGCTATATCTGTTGTTGATTCAGCAG 535  
OY 157 GAATGAG 165  
DB 536 AAGGTTAG 544

## RESULT 4

V89065  
ID V89065 standard; cDNA; 414 BP.

AC V89065.  
DT 15-FEB-1999 (first entry)  
DE EST clone B253.

KW Human: secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibin; chemokinesis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
KW gene therapy; ss.  
OS Homo sapiens.  
PN WO9845436-A2.  
PD 15-OCT-1998.  
PR 10-APR-1998; U06955.  
PR 10-APR-1997; US-838821.

PA (GENY ) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racie LA, Spaulding V, Treacy M;  
DR WPI: 99-070077/06.  
PT New polynucleotides encoding human secreted proteins - derived from

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.  
PS Claim 1; Page 99; 61pp; English.

CC The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
SQ Sequence 414 BP; 98 A; 90 C; 100 G; 126 T;

Query Match 11.3%; Score 28.6; DB 1; Length 414;  
Best Local Similarity 55.6%; Pred. No. 2.4;  
Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 47 AGCTCAGAGCTAGAGAGCTGTCAAATGACAGAGGTTTACAGACATATGCTTCAG 106  
DB 224 AGTTTGAATCCAGATATGCTGTAAGACAGGACCTTCTGCTGATTTCCACACAG 283  
OY 107 TCAAGTTCAGACAAATGCTTCATGATGCTTTGGGGC 145  
DB 284 GAATTTATACAGAGTGCTGACATGCTGCTGAGGCC 322



```

RESULT 5
X24753/c
ID X24753 standard; cDNA: 7812 BP.
AC X24753:
DT 21-JUN-1999 (first entry)
DE Human Interleukin-18 binding protein gene.
KW Interleukin-18 binding protein; IL-18BP; IL-18BPb; splice variant;
human; autoimmune disease; inflammation; diabetes; pancreatitis;
rheumatoid arthritis; graft rejection; sepsis; chronic hepatitis;
psoriasis; inflammatory bowel disease; multiple sclerosis;
KW Ischemic heart disease; Ischemic brain injury; gene therapy; ss.
OS Homo sapiens.
PN WO9909063-A1.
PD 25-FEB-1999.
PF 13-AUG-1998: IL0379.
PR 22-JUL-1998: IL-125463.
PR 14-AUG-1997: IL-121554.
PR 27-AUG-1997: IL-121639.
PR 29-SEP-1997: IL-121860.
PR 06-NOV-1997: IL-122134.
PA (YEDA ) YEDA RES & DEV CO LTD.
PI Dinarello C, Kim SH, Novick D, Rubinstein M;
DR WPI: 99-180975/15.
PT New interleukin-18 binding protein - useful for treating human
PT diseases, including autoimmune disease and inflammation
PS Disclosure: Page 60-63: 100pp; English.
CC The present sequence comprises human interleukin-18 binding protein
CC (IL-18BP) genomic DNA. It was produced from 5 genomic clones
CC isolated from a human genomic library by screening with full-length
CC human IL-18BP cDNA. No exon coding for a transmembrane receptor
CC was identified within the 7.8 kb sequence. 4 Splice variant clones
CC (see X24749-52) were identified that code for putative soluble
CC secreted proteins, designated IL-18BP-a-d (see W8004-007). The
CC IL-18BP locus contains an additional gene, coding for the nuclear
CC mitotic apparatus protein 1 (NMN1), positioned on the minus strand.
CC This localised the IL-18BP gene to human chromosome 11q13(36).
CC IL-18BP polypeptides capable of binding IL-18 and/or modulating
CC and/or blocking IL-18 activity are provided by the invention.
CC Methods for their isolation and recombinant production, DNA vectors
CC expressing them, vectors useful for their expression in humans and
CC other mammals, and antibodies against them are also provided.
CC IL-18BP polypeptides, and DNA encoding them, can be used to treat
CC conditions requiring the protein (claimed). Such conditions
CC include autoimmune diseases, type I diabetes, rheumatoid arthritis,
CC graft rejections, inflammatory bowel disease, sepsis, multiple
CC sclerosis, ischemic heart diseases, ischemic brain injury, chronic
CC hepatitis, psoriasis, and chronic/acute pancreatitis. IL-18BP is
CC also useful for purifying IL-18 (claimed).
SQ Sequence 7812 BP; 1689 A; 2278 C; 2167 G; 1678 T;

Query Match 11.3%; Score 28.6; DB 1; Length 7812;
Best Local Similarity 57.1%; Pred. No. 7.9;
Matches 52; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 131 TGAATGCTTTGGGCGCTTCATCAGAAATGAGCTGTCTCAGCGCTCATCCAGCT 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1946 TCGTGAAGTTGGGAGCGCGCTCAGGCCACAGAGATGATGCTTCATCAGG 1887
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 191 TGAGAGTAAGGATTAACCTCCAGAACAGC 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1886 ACACAGCGACAGTACGCTCTTCAGAAAGAC 1856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
ID 005733/c
AC 005733:
DT 07-JAN-1991 (first entry)
DE Glutathione-S-transferase (GST)-7 clone insert.
KW Glutathione-S-transferase (GST)-7; liver fluke infection vaccine;
KW ss.

```

```

OS Fasciola hepatica.
FH key Location/Qualifiers
FT poly_site 745..745
FT /tag=a
FT PN WO99008819-A.
PD 09-AUG-1990.
PE 31-JAN-1990; AU0027.
PR 31-JAN-1989; AU-002481.
PA (DARA-) DARATECH PTY LTD.
PI Crameri SL, Panaccio M, Wilson R, Wijffels GL, Splithill TW,
PI Campbell NJ, Thompson C, Sexton JL;
DR WPI: 90-260928/34.
DR P-PSDB: R06499.
PT Vaccine against liver fluke infection in ruminants. esp. sheep -
PT comprises glutathione-S-transferase extracted from adult fasciola
PT hepatica worms.
PS Claim 9; Fig 15; 44pp; English.
CC A cDNA library was constructed from total RNA of adult worms of the
CC Compton strain of Fasciola hepatica in either lambda g11 or lambda
CC ZAP phage vectors. The library was screened with rabbit antiserum
CC raised to purified GSTs of F. hepatica. Clone GST-7 was isolated from
CC a lambda ZAP library and was used to isolate other homologous
CC sequences. It contains a poly(A) tail indicating the 3'
CC end of mRNA has been cloned. The protein derived from this sequence
CC begins 7 amino acids from the N-terminus of GST(FH26a) and is homo-
CC logous to it.
CC This sequence can be used in a vaccine for preventative treatment
CC of liver fluke in ruminants.
CC See also 005732, 005734-005736 and R06503-R06506.
SQ Sequence 753 BP; 219 A; 146 C; 176 G; 212 T;

Query Match 11.1%; Score 28.2; DB 1; Length 753;
Best Local Similarity 55.7%; Pred. No. 4.1;
Matches 54; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 83 AGCTTACAGACATATGCTTCAGATCATCAGAAATGGCCATGATGCTTTGG 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 675 AGATTACAGGAATCTCTCATGATGATGACGACAGATGCTCATCAGCAGTGG 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 143 GGCCTTTCATCAGGAATGAGACTGTCTTCAGGCC 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 615 AGCGTCTCCACACCAAAAGAGAGAGTCCACGAGTGC 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
ID T69307 standard; cDNA: 371 BP.
AC T69307:
DT 26-FEB-1998 (first entry)
DE Murine metastatic nucleic acid sequence.
KW Mouse; murine; tumour; cancer; metastatic sequence; detection;
KW diagnosis; treatment; metastasis; hyperplasia; dysplasia;
KW hypertrophy; screening; ss.
OS Mus musculus.
PN WO9718454-A2.
PD 22-MAY-1997.
PE 15-NOV-1996; U18567.
PR 30-JAN-1996; US-584031.
PR 16-NOV-1995; US-006838.
PA (THOM/) THOMPSON T.
PI Thompson T;
DR WPI: 97-289397/26.
PT Identifying tumour metastatic sequences - by introducing transfected
PT cells into host mammal and analysing primary and metastatic
PT sequences by differential display PCR
PS Disclosure: Fig 12BY; 102pp; English.
CC Mouse Urogenital Sinus (UGS) tissue was isolated from 17 day old
CC mouse embryos. The UGS cells were infected with retroviruses,
CC cultured and implanted under the renal capsule of mice.
CC Reconstructions were harvested 5 weeks later, when they showed
CC signs of distress from the tumour burden. Metastatic tumours were
CC isolated from a site outside the renal capsule. RNA was isolated

```

CC from primary tumours and metastases, reverse transcribed and  
CC subjected to differential display PCR. The sequences were analysed  
CC to obtain metastatic sequences, e.g. the present sequence. The  
CC method can be used to detect, diagnose and treat disorders related  
CC to metastasis, or treat malignant or non-malignant disorders, e.g.  
CC hyperplasia, dysplasia and hypertrophy. The metastatic sequence can  
CC be used to screen a biological sample for metastasis, and it or its  
CC expression product may also be used to treat a metastatic disorder.  
SQ Sequence 371 BP; 103 A; 75 C; 84 G; 109 T;

Query Match 11.1%; Score 28; DB 1; Length 371;  
Best Local Similarity 55.0%; Pred. No. 3.6;  
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 72 AATATGACAGAGTTTACAGACATATGCTTCAGATCAGAAATGGCCCTCAT 131  
DB 75 AATATGCTCAGATGTTACACAGCTGTGAACAGAGGCTCAGATCGCTGATGGCTTG 134  
OY 132 GATGCTTTGGGCGCCCTTCATCAGGAATGAGCTCT 171  
DB 135 AGAATGTTTCTTGTGACATCAGAACTGAAATGTTT 174

## RESULT 8

DB 135 AGAATGTTTCTTGTGACATCAGAACTGAAATGTTT 174

DE 01-MAR-1999 (first entry)  
DE Kidney injury associated molecule HM036 cDNA clone.  
KM Kidney injury associated molecule: kidney injury related molecule;  
KM Kidney tissue growth promotion; regeneration; renal condition;  
KM acute renal failure; acute nephritis; tumour; ds.  
OS Rattus sp.

FT Key Location/Qualifiers  
FT misc-feature complement (2541..2575)  
FT /tag= a  
FT /label= SMC\_24406

PN 26-NOV-1998

PF 22-MAY-1998; U10547.

PR 23-MAY-1997; US-047491.

PR 23-MAY-1997; US-047490.

PA (BIOJ) BIOGEN INC.

PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H:

DR MPI: 99-045312/04.

PT Kidney injury associated molecule, KIM, polypeptides - upregulated  
PT in injured or regenerating tissues, useful to promote tissue growth  
PT and regeneration, especially to treat renal conditions

PS Claim 9; Page 70-71; 213pp; English.

CC The present sequence represents a kidney injury associated molecule  
CC (KIM) cDNA clone. KIM proteins can be administered therapeutically  
CC by expressing KIM encoding polynucleotides, to promote growth and/or  
CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins  
CC are upregulated in injured or regenerating (especially renal) tissues.

CC KIM fusion proteins, conjugates, antibodies and vectors can also be used  
CC therapeutically, e.g. these or the KIM proteins may be included with an  
CC acceptable carrier in pharmaceutical compositions, useful for therapy/  
CC prophylaxis of conditions associated with dysfunction/dysregulation of  
CC KIM genes or proteins, especially renal diseases or impairments of renal  
CC function in humans (e.g. acute renal failure, acute nephritis). The  
CC polynucleotides can be used to produce antisense sequences which, when  
CC internalised into cells, can disrupt expression of a cellular KIM gene,  
CC also useful in therapy (e.g. to block the growth of tumours dependent on  
CC KIM for growth) or compositions. The proteins and polynucleotides are  
CC useful diagnostically e.g. to detect and quantify renal injury/disease  
CC (indicative of increased risk, or presence of, renal injury or impaired  
CC function), or abnormal responses to tissue injury (indicative of  
CC increased risk, or presence of, an autoimmune response or abnormal  
CC tissue growth arising from/affecting renal tissue). The proteins can  
CC also be used to locate KIM-producing cells (especially specific loci,  
CC e.g. tissue masses abnormally producing/expressing KIM such as tumours  
CC arising from/affecting renal tissue), by contacting cells with an

CC imageable KIM-binding reagent and imaging reagent accumulation.  
SQ Sequence 3531 BP; 846 A; 795 C; 732 G; 1158 T;

Query Match 11.1%; Score 28; DB 1; Length 3531;  
Best Local Similarity 55.0%; Pred. No. 9;  
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 30 GCTTGGGCGCCCTTCAGACCTCAGAGAGAGAGGCTGCCAATGACAGAGGTTTA 89  
DB 2738 GCCGTGGGACTCAGCTGTTTCCACCTGGAGCTCTGACTGCTGTAACATGTATA 2797  
OY 90 CAGACATATGCTTCAGATCAGATCAGAAATGGCCCTCA 129  
DB 2798 TACAATATCCCTCGATCCTCTTAGCATACACGAGA 2837

## RESULT 9

DB 2798 TACAATATCCCTCGATCCTCTTAGCATACACGAGA 2837

DE 16-MAR-1994 (first entry)  
DE Human brain Expressed Sequence Tag EST01961.  
KM Gene transcription product; genetic markers; tagging; in vivo;  
KM transcription; mapping; locations; chromosomes; chromosomal; ss.  
OS Homo sapiens.

PN WO9316178-A.

PF 19-AUG-1993.

PF 12-FEB-1993; U01294.

PR 12-FEB-1992; US-837195.

PA (USSH) US DEPT HEALTH & HUMAN SERVICE.

PI Adams MD, Moreno RF, Venter CJ;

DR MPI: 93-272882/34.

PT Enriched oligonucleotides and corresp. sequences - used as  
PT markers for human genes transcribed in-vivo, facilitate tagging  
PT of most human genes

PS Example 4; Page 262; 500pp; English.

CC The Expressed Sequence Tag was isolated from a human brain cDNA  
CC library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes  
CC on chromosomes, for individual or forensic identification, for mapping  
CC locations of disease-associated genes, for identification of tissue  
CC type, and for prep. of antisense sequences, probes and constructs.

CC EST01961 has been putatively identified as encoding a Notch/Xotch  
CC protein. See also 059041-061440.

CC Sequence 302 BP; 72 A; 79 C; 69 G; 77 T;

Query Match 11.0%; Score 27.8; DB 1; Length 302;  
Best Local Similarity 47.7%; Pred. No. 3.8;  
Matches 74; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY 11 TGCATCATCCACACAGTCGCTTGGGCGCCCTGCAGCTCAAGACTAGAGAGCTGC 70

DB 65 TGAGACCTCCAGTACAGTTTCCATGATGACAGGATTCMCAGCACTTCTTCACTGTTN 124

OY 71 CAAATGACAGAGAGTTTACAGACATATGCTTCAGATCAAGTTTCAGAACATGGCCCTCAT 130

DB 125 AGTAGCAGCTGGGGGTGATGGGGTCCCTCGGGCATATACAGCGGAACCATTCACACGT 184

OY 131 TGATGCTTTGGGCGCCCTTCATCAGGAATGGAG 165

DB 185 TGATACATGTGACACCTTCGACAGGGATTGGNG 219

## RESULT 10

V21209\_03/c

Continuation (4 of 17) of V21209 from base 300001 (Methanococcus jannaschii circular  
WP Sequence Split into 17 fragments LOCUS V21209 Accession V21209  
WP Fragment Name Begin End  
WP V21209\_00 1 110000  
WP V21209\_01 100001 210000



[illegible]

Query Match	Best Local Similarity	Score	DB	Length
Matches 78; Conservative 0; Mismatches 84; Indels 0; Gaps 0;	10.9%;	27.6;	DB 1;	3872;
46 CAGCTCAAGAACTAGAGAGACCTGTCGCAAAATGACAGAGAGTTACAGACATATGCTTCAG 105	48.1%;	Pred. No. 13;		
2213 CACATCAAGGCGTGAAGATAGTTGGCAAAATTCACAGTACCTTGATTTATTTATCTTTG 2154	0;			
106 ATCAAGTTGGAACATATGCGCTTCATTGATCTTTGGGCGCTTTCATCAGGAATGAG 165				
2153 TGAATAAGTGAATAAAGTACGTTATATATCTGTGTACATTAAGATGATGAAT 2094				
166 TGTCTCTCAGCGCTCCATCCAGGTTGAGAGTAAAGGATTA 207				
2093 CATTCACTCAATGATCTGTACATGCAAAAACACAGATATTAA 2052				
RESULT 14				
ID X13032/c				
AC X13032; standard; DNA; 8726 BP.				
DT 19-MAR-1999 (first entry)				
DE Enterococcus faecalis genome contig SEQ ID NO:95.				
KW Enterococcus faecalis; contig; detection; Enterococcal infection;				
OS vaccine; attenuation; computer readable medium; ds.				
PN WO980555-A2.				
PD 12-NOV-1998.				
PE 04-MAY-1998; U08985.				
PR 14-NOV-1997; US-066009.				
PR 06-MAY-1997; US-044031.				
PR 16-MAY-1997; US-046655.				
PA (HUMA-) HDUMAN GENOME SCI INC.				
PI Barash SC, Dillon PJ, Kunsch CA;				
DR WPI: 99-045171/04.				
PT New isolated Enterococcus faecalis polynucleotides and polypeptides				
PT - used to develop products for the detection of Enterococcus and for				
PT use in vaccines for prevention or attenuation of Enterococcus				
PT infection.				
PS Claim 1; Page 626-631; 2084PP; English.				
CC A computer readable medium has been developed which has recorded on it				
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.				
CC X12938 to X13919 represent these nucleotide sequences which are primary				
CC nucleotide sequences, also known as contigs. The computer-based system				
CC can identify fragments of the Enterococcus faecalis genome with				
CC commercial importance. The products can be used to detect the presence				
CC of Enterococcus faecalis in samples. They can also be used for				
CC diagnosing Enterococcal infection in an animal and monitoring				
CC progression of disease, and for identifying agents which can be used to				
CC modulate the growth or pathogenicity of Enterococcus faecalis, or				
CC another related organism, in vivo or in vitro. In particular the				
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences				
CC can be used in vaccines to prevent or attenuate an Enterococcal				
CC infection.				
SO Sequence 8726 BP; 2834 A; 1460 C; 1912 G; 2509 T;				
Query Match	10.8%;	Score 27.4;	DB 1;	Length 8726;
Best Local Similarity	50.4%;	Pred. No. 21;		
Matches 67; Conservative 0; Mismatches 66; Indels 0; Gaps 0;				
55 AACTGAGAGAGCTGTCGCAAAATGACAGAGGTTTACAACATATGCTTCAGATCAAGTTC 114				
8540 AACCAATATTCATTAATAAAGTTTCAAGATATGTCATCACCATTATCTTATATATCG 8481				
115 AGAACAAATGGCCCTCATGATGCTTTTGGGCGCTTTCATCAGGAATGAGAGTGTCTGC 174				

```

Db 8480 CTTAAATATGCTCTGACACATGTGGTGTGATGCATAAAGCAACGAATATATC 8421
QY 175 AGCGCTCATCCA 187
Db 8420 AGCGTATTGCA 8408

```

## RESULT 15

```

T30736
ID T30736 standard; DNA; 2457 BP.
AC T30736;
DE 20-OCT-1996 (first entry)
DE Rat cryptdin 1 gene.
KW Cryptdin 1; antibiotic; antimicrobial; defensin; inflammation;
KW antiinflammatory; inflammatory bowel disease; pancreatitis;
OS Rattus sp.
FH Key
FT Location/Qualifiers
FT exon 1..1390
FT /tag- a
FT /codon_start- 1218..1221
FT /note- "exon 1 codes for the 5' untranslated
FT region and cryptdin-1 prepro sequence"
FT 1391..1972
FT /tag- b
FT 1973..2457
FT /tag- c
FT /note- "exon 2 codes for cryptdin-1 and
FT 3' untranslated region"

```

```

PN MO9616075-A1.
PD 30-MAR-1996.
PF 05-OCT-1995; U13328.
PR 18-NOV-1994; US-342268.
PA (SHRI-) SHRINER'S HOSPITAL FOR CRIPPLED CHILDREN.
PA (REGC ) UNIV CALIFORNIA.
PI Ouellette AJ, Seisted ME;
DR WPI: 96-268527/27.
DR P-PSDB; R98786;
DR P-PSDB; R98791.
PT New isolated cryptdin peptide(s) - which have antimicrobial
PT activity, used partic. in the detection and treatment of
PT inflammatory pathologies
PS Claim 31: Page 63-64; 103pp; English.
CC The rat cryptdin-1 gene (T30736) codes for the precursor
CC (R98786) of cryptdin-1 (R98791), an antimicrobial peptide
CC that exhibits activity against a broad range of intestinal and
CC opportunistic pathogens. Rat cryptdin-1, -2, and -3 genes
CC (see also T30737-38) were isolated from a Sprague-Dawley
CC genomic library cloned in EMBL3. The cryptdin-1 gene, or
CC corresponding cDNA (see also T30733), can be used to produce
CC large amounts of cryptdin-1 for use in treating inflammatory
CC pathologies of the intestine.
SQ Sequence 2457 BP; 700 A; 526 C; 536 G; 695 T;

```

```

Query Match 10.7%; Score 27; DB 1; Length 2457;
Best Local Similarity 49.6%; Pred. No. 17;
Matches 69; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```

```

QY 93 ACATATGCTTCAGATCAAGTTCAGAACATGGCTCATTCATGCTTTGGGGCCCTTCA 152
Db 1193 ACCAATCCCTCCAGTGACTTCAGCCATGAAGACTCTTGTCTCTCTGCCCCCTTCC 1252
QY 153 TCAGAAATGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAAGTAAGGATTAACTTC 212
Db 1253 TCGTGCATTCAGCTCCAGCTCCAGCTATCCATTAAGAGCAGAGAAGACTTAAACTG 1312
QY 213 CAGAACGCCAGTGATGA 231
Db 1313 AGGAGCAGCCAGCAGATGA 1331

```

Search completed: August 6, 1999, 04:03:09  
 Job time: 1341 sec

This Page Blank (uspto)

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:39:19 ; Search time 3268.17 Seconds

(without alignments)  
185.865 Million cell updates/sec

Title: US-09-049-696-11

Perfect score: 191  
Sequence: 1 GGGTTTGTAGTGGACCAAAA.....AACAAAGACACCAGCAAAATT 191

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hlg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hlg1:\*  
35: gb\_hlg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	191	100.0	3311	11	AF039400	AF039400 Homo sapi

2	191	100.0	2826	42	AF127036	AF127036 Homo sapi
3	133	69.6	35278	11	AF039401	AF039401 Homo sapi
4	19	9.9	13246	1	KPNPDL	M32613 K.pneumonia
5	9	9.9	83549	11	AC004752	AC004752 Homo sapi
6	5	9.4	14168	2	AE001281	AE001281 Chlamydia
7	8	9.4	122322	7	ATP6118	AT022198 Arabidops
8	18	9.4	142928	9	AP000031	AP000031 Homo sapi
9	18	9.4	100000	9	AP000134	AP000134 Homo sapi
10	18	9.4	100000	9	AP000212	AP000212 Homo sapi
11	11	9.4	67008	9	AP000252	AP000252 Homo sapi
12	18	9.4	176153	11	AC006287	AC006287 Homo sapi
13	18	9.4	205573	34	CNS00004	AT049778 Homo sapi
14	17	8.9	38916	1	MICB2548	AL023093 Mycobacte
15	17	8.9	91149	8	AF058919	AF058919 Arabidops
16	17	8.9	108484	8	ATAC004238	AC004238 Arabidops
17	17	8.9	57550	8	ATAC007212	AC007212 Arabidops
18	17	8.9	73511	9	HS65A6	Z92546 Human DNA s
19	19	8.9	121780	9	HS774124	AL031290 Human DNA
20	20	8.9	69017	10	HS41018	AL031732 Human DNA
21	17	8.9	78293	11	AC004636	AC004636 Homo sapi
22	17	8.9	158078	11	AC005094	AC005094 Homo sapi
23	17	8.9	179651	11	AC005553	AC005553 Homo sapi
24	17	8.9	175818	11	AC006257	AC006257 Homo sapi
25	17	8.9	159424	11	AF064865	AF064865 Homo sapi
26	17	8.9	189134	11	HOAC004682	AC004682 Homo sapi
27	17	8.9	128026	12	AC004155	AC004155 Mus muscu
28	28	8.9	335	14	G12349	G12349 Nile tilapi
29	29	8.9	417	14	G25428	G25428 human STS E
30	30	8.9	417	14	G27691	G27691 human STS S
31	17	8.9	183352	34	AC006459	AC006459 Homo sapi
32	17	8.9	27859	34	AC006629	AC006629 Caenorhab
33	17	8.9	215917	34	AC006764	AC006764 Caenorhab
34	17	8.9	40734	34	HS303A1	AL031704 Homo sapi
35	17	8.9	166597	34	HS467L1	Z98884 Homo sapien
36	17	8.9	110000	34	HSDJ689N3_1	Continuation (2 of
37	17	8.9	146126	34	HSDJ97781	AL050318 Homo sapi
38	17	8.9	193003	35	AC007464	AC007464 Homo sapi
39	17	8.9	34216	36	CEL80478	U57054 Caenorhabdi
40	16	8.4	7499	1	ECHSDMSR	X13145 Escherichia
41	16	8.4	1824	1	PPDEHRI	Z48919 P.putida de
42	16	8.4	1824	1	PPU23716	U23716 Pseudomonas
43	16	8.4	279110	1	RPRX03	ALJ35272 Rickettsi
44	16	8.4	407	1	RZ82753	Z82753 R.prowazeki
45	16	8.4	121803	42	AC004835	AC004835 Homo sapi

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS	AF039400					
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA, complete cds.					
ACCESSION	AF039400					
NID	94009457					
VERSION	AF039400.1 GI:4009457					
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
AUTHORS	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.					
TITLE	Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins					
JOURNAL	Genomics 54 (2), 200-214 (1998)					
MEDLINE	99047526					
REFERENCE	2 (bases 1 to 3311)					
AUTHORS	Gruber,A.D., Elble,R. and Pauli,B.U.					
TITLE	Direct Submission					

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

REFERENCE  
1 (bases 1 to 2826)  
Agnel M., Vermet, T. and Culouscou, J.-M.  
Cloning of three human homologs of bovine epithelial chloride  
channel

JOURNAL  
REFERENCE  
2 (bases 1 to 2826)  
Agnel M. and Culouscou, J.-M.  
Direct Submission  
Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrieres, Rueil-Malmaison 92500, France

## FEATURES

source

location/Qualifiers

1. .3311

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="1p22-p31"

/tissue\_type="small intestine"

1. .3311

/gene="hCLCA1"

352. .3096

/note="transmembrane glycoprotein"

/codon\_start=1

/product="calcium-dependent chloride channel-1"

/protein\_id="AAC95428.1"

/db\_xref="PID:94009458"

/db\_xref="GI:4009458"

/translation="MGPFSSVFILILHLEGALNSLIQINNGYEGIVAIIDPNP  
EDFLIOIKMVTQASLYFEATGRKFRFNVAIIPETWKTRADYVRPLETKNA  
DYLAEETPGNDEPTLOMGNGEGERHILTPDLAKKLAIEYGPGRFAVEHMA  
LRMGVDEYNDDEKFLYLSNGRIQAVRCASGITGNVVKCGGSCYTKRCTFNVTGL  
YKGCGEVLOSROTEKASIMFAOHVDSIVECTEONHNKRNKONOKNIRSTWETV  
RDSDEKRTKPTMPTOPNPPTFSILOQRVCLVLDKSGMATGNRLRLNOAGOLF  
LQVLEGSVMGVTEDSAHVOSELIQINSQSDRDLAKRLPAASGGTSICGLBSA  
FTVIRKRYPTDGESEIVLTDGEDNTISGFENEKOSGAIITVVALPSAOEDELISK  
MTGGLQIYASDQVQNNGLIDAFGLSSGNGAVSORSIOLESKGLLQNSQMMNGTVY  
DSTVGRDLEFLITWITOPQIILMDPSGKOGGFVNDTMMAYLQIPGIKAVGTWY  
SLAASSQTLTLVTSRASNAATLPITVTSKNTKDSKPSPLVYVYANIRKVASPILNA  
SVTALIESVNGKTVLELDNGAGADATKDGVSRYETTYDNGRYSVKRALGYN  
AARRVITPOOSGALYIPGNIENDEIOMNPREINKDVOHKOVCFSRTSSGGSFYAS  
DVFNAPITPDLFPPGQITLDAEIHGSLINLTWITAGDDYDHGTAKYIIRISTSLD  
LRKFENESLOVNTTALIPKEANSEVEFLKPNITPENGTDLFLIAOAVDVKLSEI  
SNIRAVSLFIPQTPETPSPDTSAPCPNHNISTIPGIIHLIKIMKMWIGELQSLIA

BASE COUNT 1028 a 692 c 742 g 849 t

ORIGIN

Query Match 100.0%; Score 191; DB 11; Length 3311;

Best Local Similarity 100.0%; Pred. No. 6e-99;

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1972 GGCTTTAGTGGACAAAACACAAAATGGCTTACCTCCAAATCCAGGATTGCTAAG 60

OY 1 GGCTTTAGTGGACAAAACACAAAATGGCTTACCTCCAAATCCAGGATTGCTAAG 60

Db 1972 GGCTTTAGTGGACAAAACACAAAATGGCTTACCTCCAAATCCAGGATTGCTAAG 2031

OY 61 GTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTTCAAAACCTTGACCTGTCACG 120

Db 2032 GTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTTCAAAACCTTGACCTGTCACG 2091

OY 121 TCCCGTGGTGCATGCAATGCTGCTCCAAATAGAGTCTCCAAAGCAAGAGAC 180

Db 2092 TCCCGTGGTGCATGCAATGCTGCTCCAAATAGAGTCTCCAAAGCAAGAGAC 2151

OY 181 ACCAGCAAAAT 191

Db 2152 ACCAGCAAAAT 2162

RESULT 2

AF127036 2826 bp mRNA PRI 16-APR-1999

LOCUS AF127036

DEFINITION Homo sapiens calcium-activated chloride channel protein 1 (CaCC1)

ACCESSION AF127036

VERSION 94585468

KEYWORDS AF127036.1 GI:4585468

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

BASE COUNT 875 a 623 c 632 g 696 t

ORIGIN

Query Match 100.0%; Score 191; DB 42; Length 2826;

Best Local Similarity 100.0%; Pred. No. 6e-99;

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1625 GGCTTTAGTGGACAAAACACAAAATGGCTTACCTCCAAATCCAGGATTGCTAAG 1684

OY 61 GTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTTCAAAACCTTGACCTGTCACG 120

Db 1685 GTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTTCAAAACCTTGACCTGTCACG 1744

OY 121 TCCCGTGGTGCATGCTGCTCCAAATAGAGTCTCCAAAGCAAGAGAC 180

Db 1745 TCCCGTGGTGCATGCTGCTCCAAATAGAGTCTCCAAAGCAAGAGAC 1804

OY 181 ACCAGCAAAAT 191

Db 1805 ACCAGCAAAAT 1815

RESULT 3

AF039401 35278 bp DNA PRI 15-DEC-1998

LOCUS AF039401

DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,

complete cds.



ACCESSION AF039401  
MDB 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreue,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human CLCA1, the first human member of the family of  
Ca2+-activated Cl--channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
Source 1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="11 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,  
9696..9801,15454..15631,18591..18809,19783..20010,  
22253..22427,24522..24628,26644..26859,27447..27708,28766,  
28936..28936,31840..32079,32919..33310)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,  
15454..15631,18591..18809,19783..20010,22253..22427,  
24522..24628,26644..26859,27447..27708,28766..28936,  
31840..32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPRSSVFLLILHLLEGALSNSLIOLNNNGEIVVAIDPNV  
EDERLIOIKMVTQASLXLEPATGRKRFKVALIIPETKTRADYVRPKETKNA  
DVLVAESTPGRNDEPFTEDMGNGEGERIRHLPFLIGKLAIEFGPKAFVHEMAH  
LRMGVFDEYNDEKFTLSNGRIQAVRCAGSAGITGVNVRKCGSCYTRCFNKVTGL  
YEKGEFVLQSGKOTKASIMFAQHYDSIVEFTEONHREKAPNKQNCNLSTWEVI  
RDEDEFKTTPMTQTPNPFTSLQIGRIYCLVATGSGMSGRRLNRLNAGQLFL  
LQTVELGSMGVNTEDSAHOSELIOINSGSDRPTLAKRLPAASGGSTISCSGLRSA  
FTVIRKRYPTDGSSEIVLTDGEDNTISCFENVKSGAIHTHVALGPRAADEBLK  
MTGGLQITASDOVONNGLIDAFGLSNGNAVSQSIDLESGLTQNSOMNGVIV  
DSYVGKDTFLITWTQPPQILLMDPSQKGGFVNDNTKAAVQIIGIKAVGWKY  
SLQASQKDTLLTVSRASNATLPTITVSKTKDTSKPPSLVYVANAIRQASPLIRA  
SVTALIESVNGKTVTLELDNGAGADATKDDVRYRTYDTNGRYSVKYALGYN  
AARRVITPOOSGALYIPGMIENDEIOMPPREIKDDVQHKOVCFESTSGGSPVAS  
DVNAPRIPDLRPPQITDLNAEIHGSLINTLTWTPAGDDYDHGTAKHTIIRISITLD

LRDKNESIOVNTATLIPKANSSEFVLPFRKENTFENGCTDLFTIAQAVDKLSEI  
SNIRVSELPFPQIPPEPSPDETSARCPNHHNSTFGIHILIMKMKHIELOLSTIA  
6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 69.6%; Score 133; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 7,2e-66;  
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 AGGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCTGACTGTCA 118  
|||||  
DB 27445 AGGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCTGACTGTCA 27504  
QY 119 GCTCCGCGTGCCTCAATGCTACCTGCTCCCAATTACAGTACTTCCAAAAGCAAGG 178  
|||||  
DB 27505 GCTCCGCGCGCTCAATGCTACCTGCTCCCAATTACAGTACTTCCAAAAGCAAGG 27564  
QY 179 ACACGAGCAATT 191  
|||||  
DB 27565 ACACGAGCAATT 27577  
RESULT 4  
KPNPUL KPNPUL  
LOCUS KPNPUL 13246 bp DNA BCT 15-JUN-1990  
DEFINITION K.pneumoniae pululanase secretion operon genes pulC, pulD, pulE,  
pulf, pulG, pulH, pullI, pullJ, pulK, pulL, pulM, pulN, pulO, and pulO.  
complete cds.  
ACCESSION M32613 M24118 X14954  
NID 9149301  
VERSION M32613.1 GI:149301  
KEYWORDS envelope-associated protein; pululanase secretion protein.

SOURCE	K.pneumoniae (strain UNF5023) DNA.
ORGANISM	Klebsiella pneumoniae
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Klebsiella.
AUTHORS	1 (sites)
TITLE	d'Enfert,C., Reys,I., Wandersman,C. and Pugsley,A.P.
	Protein secretion by gram-negative bacteria: Characterization of two membrane proteins required for pullulanase secretion by Escherichia coli K-12
JOURNAL	J Biol. Chem. 264, 17462-17468 (1989)
MEDLINE	90008916
REFERENCE	2 (bases 1 to 13246)
AUTHORS	Pugsley,A.P., D'Enfert,C., Reys,I. and Wandersman,C.
JOURNAL	Unpublished (1990)
MEDLINE	3 (sites)
REFERENCE	Reys,I. and Pugsley,A.P.
AUTHORS	Five additional genes in the pulC-O operon of the gram-negative bacterium Klebsiella oxytoca un55023 which are required for pullulanase secretion
TITLE	Mol. Gen. Genet. 222, 176-184 (1990)
JOURNAL	91109698
MEDLINE	4 (bases 1 to 13246)
REFERENCE	Posset,O., d'Enfert,C., Reys,I. and Pugsley,A.P.
AUTHORS	Pullulanase secretion in Escherichia coli K-12 requires a cytoplasmic protein and a putative polytopic cytoplasmic membrane protein
TITLE	Mol. Microbiol. 6, 95-105 (1992)
JOURNAL	92149318
MEDLINE	[1] sites.
REFERENCE	Draft entry and computer-readable sequence for [1] kindly submitted by Pugsley,A.P., 06-MAR-1990.
FEATURES	Location/Qualifiers
source	1. 13246
	/organism="Klebsiella pneumoniae"
	/db_xref="taxon:573"
	408..1265
	/note="pulC"
	/codon_start=1
	/transl_table=11
	/protein_id="AA25125.1"
	/db_xref="PID:g149302"
	/db_xref="GI:149302"
	/translation="MSKGKIMHNSVWRLTPPKKIINAPHYIVTSIIIFPICQOLAOLTKIILPVNFTDAALSSADMTSPASASATLPTLFLGLAKTISAPGGNLDQAPYSAIRLAVTGLASTDSRAIAIMKGNQOVSGIGDNTPGGEAKIIASPRLLVYNRGRNEALPLENDPAPVAGKNSAPPARHLQDELGAQONLTHYLNISPVVNDRLSGYRLNPKGDPALFROSLRENDLAIALNGIDLDKKQARQVLAQLPELTETITLTERDGQKNDIYALARDE"
	1266..3248
	/note="puld (ttg start codon)"
	/codon_start=1
	/transl_table=11
	/protein_id="AA25126.1"
	/db_xref="PID:g149303"
	/db_xref="GI:149303"
	/translation="LIIANVIRSFSLTILFALLPRPAAAEPSASPKGDIOEFTNYSKINKTVIITDPSRGITIVRSYDMLNEBOYTFPFLSVLDYGFANYNANNGVLYVSKDKKTAIVPASPADAPGIDEVVTRVPTLNAAARDLAPLRQLNDNAGVSVAHYEPSNVLLMTGKRAVTKRLTIVERYDNAGDSVTVPLSNASADVVKLTETLNKTKSKSALDSEVAVNADERTNAVLVSEGESEKRIIAIMKQIDROQATGQGVLYIKYAAASDVEVLVIGTISSTMOSEKQAKPAVALDKNIIIIKAGQTNALITVAADVANDDERYIAQDIDIRPOVIVEAIIAEYODAGLQNLQNMKNKAGMTOFTNSGLPISTAIAGNOUNKGTGVSLSLASSFSENGIAGFPQGNAMKILTLSSTKNDILATPSITYLDNMEITFVQGEVPEVLIGSQTSGDNTFTVERKTVGILKVPQIINSGSVLLEDEYSSVADAASSTSDLGATENTRTVNNAVLVGSEIVVGGLLDKSVSDTADVPLIGIIPVIGAFRSTSKVSKRNLMLEIRPTVIRDRDEYRQASGGQYTAFNDAOSKORKEENNDAMLNDLLEIYIPRODTAFARQVSAIDAFNIGNL"
	3249..4738
	/note="pule"
	/codon_start=1
	/transl_table=11
	/protein_id="AA25127.1"
	/db_xref="PID:g149304"
	/db_xref="GI:149304"
	/translation="MTPAERRPPLPFGYARASHVWMLSSGESCEVFLAVTAPADLI
	EEARVAMPFRLEEREEAEKLVLSYGRDSAEARRMADIGNELDYTLAEELPDLI
	DLALDESDAPITRIILNAMLTEIAEKASDIIETFERHQLREFRVDGLVEIRLPQ
	RLLADLSDIRKIVASLSDIAEKERIPQDGRMALRIGRAVRSVLPSSGSEVRLVRL
	DKSNVLDLITLGMFTALLROYDGIARPHGIVLVYTGSGSKTTLVAALSDARAR
	RNMTEDEPREVELEGIGOTONAVANDMFAAGLIIIRQDDPVIVETIDGFAO
	AVOASITGHLVLSLTNTNSALAIISRQDMGCEPFLSTSLAASQVLRILCPHC
	QOEPANADIAHQEITAPGALMOPKCAACGCTGGRGIGHELLVDRVRAIMIHRO
	ENSVTLIQDGLDGYMTLRRAGREKALGITTSQEVLRVTEQPIAAC
	4740..5885
	/note="pulf"
	/codon_start=1
	/transl_table=11
	/protein_id="AA25128.1"
	/db_xref="PID:g149305"
	/db_xref="GI:149305"
	/translation="MALFERYQALDROGKPRNGVQADSARAROLLREKGNLADIDP
	AAGGSRSEMRRTSARDLALVROLATVAAAIPLERKALDAVOQSGSPOLKTLIG
	VSKGVLEGSLAEAMRGHPCFDALYCAMVAGSGRILQAMYPIVLIVASVY
	VILLSTVVPKVIQEOPTHKQALPSTRLMAMSDMLRAGCMILLAILILLIRYL
	RGAKRLAHRLLRLPLIGRVARSVNSKARTISTINASAVPELLMNRISAVLS
	AMARKOLEASADVAREGVSLEHRALEMTDLEPPMKRYVWASERSGELMSLERADNQ
	DNRLSQIOLALSLPEPLVYVAMAGVLEIYALIIQPILOINTLMS"
	5906..6328
	/note="pulg"
	/codon_start=1
	/transl_table=11
	/protein_id="AA25129.1"
	/db_xref="PID:g149306"
	/db_xref="GI:149306"
	/translation="MQRGCEFTLEIMVYVILGVASLVVPMNGNEKADROKVVS
	DIVALEGALDMYTKLDSRYPTTEGLOLVASVSEPPARVYPEGYIRRLPDPWMS
	DYOLSPGGHGQVDFISLGPDCVPESSNDIGWITGK"
	6361..6840
	/note="puli"
	/codon_start=1
	/transl_table=11
	/protein_id="AA25130.1"
	/db_xref="PID:g149307"
	/db_xref="GI:149307"
	/translation="MLTILIMVSGNWLLAFPASRDSAAQTIAREPRAQRPVQNG
	LQIGQFQGVSVHPDKQVLELRQADPAPDDGWSGRMLPLRAGVATSGSIAGG
	KLNLAPOGEANTPDNDPVLIFPGEMTPRLTIGAPGIAFNARGESLPEPEAO"
	6837..7202
	/note="puli"
	/codon_start=1
	/transl_table=11
	/protein_id="AA25131.1"
	/db_xref="PID:g149308"
	/db_xref="GI:149308"
	/translation="MKKSGMTLIEVWVAVFALAGVAAQATLOOTROLGRMEKRT
	LASWLDNOLVOLRLERNPAPLMSSETTLQAAVSWHIRMOGVETDLPOLRALDVEVR
	HAKSDETPVSSLRVYVTS"
	7199..7795
	/note="puls"
	/codon_start=1
	/transl_table=11
	/protein_id="AA25132.1"
	/db_xref="PID:g149309"
	/db_xref="GI:149309"
	/translation="MIRSSGFTLVEMHLAIIAALSVAAVYLVQVNRADITITRK
	SGRMQALQESQMAADRSQLIIPRSRDSASLPRFGRQLDSGGAIAFSRNGWPNL
	GILPSEIOWNSYVLRGQLERLANTDODPLPGSQPTVTVLRDVRAFGLRFYASGRW
	QDEMQAQTLPGLELVTLLEPYGEIRRLFLTPGDSR"
	7792..8772
	/note="pulk"
	/codon_start=1
	/transl_table=11
	/protein_id="AA25133.1"
	/db_xref="PID:g149310"
	/db_xref="GI:149310"
	/translation="MNRHORGIALMLVLLIILAMVLAASAMTESARMYOOTATTDN



```

repeat_region      /rpt_family="MIR"      complement(10541. 10627)
                    /rpt_family="MIR"      complement(11571. 11670)
misc_feature       /note="GRAIL 2 excellent exon, frame 0"
repeat_region      /rpt_family="Alu"      complement(11763. 12085)
repeat_region      /rpt_family="Alu"      12209. 12513
repeat_region      /rpt_family="Alu"      12551. 12752
repeat_region      /rpl_family="Alu"      complement(12835. 12985)
repeat_region      /rpt_family="MER4A"    13366. 13504
repeat_region      /rpt_family="MIR"      13507. 13938
repeat_region      /rpt_family="Alu"      13589. 13627
                    /note="(TAA)13"
                    /rpt_type=tandem
repeat_region      /rpt_unit=TAA
                    13955. 14032
repeat_region      /rpl_family="MIR"      complement(14170. 15070)
                    /rpt_family="Alu"      complement(15793. 15891)
misc_feature       /note="GRAIL 2 excellent exon, frame 0"
repeat_region      /rpt_family="Alu"      complement(17186. 17533)
repeat_region      /rpt_family="Alu"      17690. 17838
repeat_region      /rpl_family="Alu"      18055. 18668
                    /rpt_family="U1"      18672. 19105
repeat_region      /rpt_family="Alu"      19238. 19536
repeat_region      /rpl_family="Alu"      19577. 19973
repeat_region      /rpt_family="U1"      complement(19996. 20313)
repeat_region      /rpt_family="Alu"      complement(20587. 20800)
repeat_region      /rpl_family="HERV16"    complement(20907. 21057)
                    /rpt_family="Alu"      complement(21084. 21156)
repeat_region      /rpt_family="HERV16"    complement(23671. 23987)
                    /rpl_family="Alu"      25314. 25536
repeat_region      /rpl_family="Alu"      complement(25670. 26578)
repeat_region      /rpt_family="MER42"    complement(27174. 27323)
repeat_region      /rpt_family="LINE2"    27422. 27759
repeat_region      /rpl_family="Alu"      complement(28441. 28535)
repeat_region      /rpt_family="L1PMA2"   28603. 29006
repeat_region      /rpt_family="Alu"      28837. 28857
                    /note="(A)21"
                    /rpt_type=tandem
repeat_region      /rpt_unit=A
                    complement(29894. 30028)
repeat_region      /rpl_family="Alu"      30721. 31343
repeat_region      /rpl_family="Alu"      31003. 31030
                    /note="(A)28"
                    /rpt_type=tandem
                    /rpt_unit=A

```

```

repeat_region      complement(31406. 31718)
                    /rpt_family="Alu"
repeat_region      /rpl_family="Alu"      32095. 32382
repeat_region      /rpl_family="Alu"      complement(32401. 32727)
repeat_region      /rpl_family="Alu"      complement(33156. 33438)
repeat_region      /rpl_family="Alu"      33764. 33900
repeat_region      /rpl_family="Alu"      complement(36030. 36364)
repeat_region      /rpl_family="Alu"      37069. 37108
                    /note="(TG)20"
                    /rpt_type=tandem
repeat_region      /rpl_unit=TG
                    37205. 37287
repeat_region      /rpl_family="Alu"      37475. 37581
repeat_region      /rpl_family="MT1H"     complement(38038. 38330)
repeat_region      /rpl_family="Alu"      complement(38593. 38908)
repeat_region      /rpl_family="Alu"      38902. 38925
                    /note="(CTC)6"
                    /rpt_type=tandem
repeat_region      /rpl_unit=CTC
                    38948. 38971
                    /note="(TRCC)6"
                    /rpt_type=tandem
repeat_region      /rpl_unit=TRCC
                    39683. 39887
repeat_region      /rpl_family="Alu"      42513. 43006
repeat_region      /rpl_family="Alu"      43326. 43375
repeat_region      /rpl_family="MIR"      complement(44646. 44779)

```

```

Query Match      9.9%; Score 19; DB 11; Length 83549;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 CCCGCTCCCAATACACT 158
Db 56985 CCGTCCTCCCAATACAGT 56967

RESULT 6
AE001281      14168 bp      DNA      BCT      02-SEP-1998
LOCUS      Chlamydia trachomatis section 8 of 87 of the complete genome.
DEFINITION      AE001281 AE001273
ACCESSION      93328454
NID      AE001281.1 GI:3328454
VERSION
KEYWORDS
SOURCE
ORGANISM      Chlamydia trachomatis.
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE      1 (bases 1 to 14168)
AUTHORS      Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusev,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis
Science 282 (5389), 754-759 (1998)
JOURNAL      2 (bases 1 to 14168)
MEDLINE      Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusev,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
Direct Submission
AUTHORS
REFERENCE
TITLE

```



AKGSLTKASLGLIRASSGOSLFGORESKYHHRIRAYAPORASVDWDEPMVLVLYM  
GCYKCKIWNINISDDBREARILIERVLEAFANROIKLSGGQOARFLARSLMOKA  
DLIIMDELFSADIMASTOMVADLOELSEKRTIVIHIDLSNWKLPDHIILNKL  
VCSSGVECLTRKEAIFQAYGCELELDTYTLKLSRKYOGSC"

gene  
/gene="ytgc"  
/note="CT069"  
10575. 11930  
10575. 11930

CDS  
/gene="ytgc"  
/codon\_start=1  
/transl\_table=11  
/product="Integral Membrane Protein"  
/protein\_id="AAC67660.1"  
/db\_xref="PID:g3328463"  
/db\_xref="GI:3328463"

gene  
/gene="ytgc"  
/note="CT070"  
11920. 12876  
11920. 12876  
/gene="ytgc"  
/codon\_start=1  
/transl\_table=11  
/product="Integral Membrane Protein"

CDS  
/gene="ytgc"  
/note="CT070"  
11920. 12876  
11920. 12876  
/gene="ytgc"  
/codon\_start=1  
/transl\_table=11  
/product="Integral Membrane Protein"

Query Match 9.48; Score 18; DB 2; Length 14168;  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 162 TTCGAAAGCAACAGGA 179  
|||||  
Db 11896 TTCGAAAGCAACAGGA 11913

RESULT 7  
ATF6118 122322 bp DNA PLN 25-MAR-1999  
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F6118 (ESSKIT project).  
ACCESSION AL022198  
NID 92980757  
VERSION AL022198.1 GI:2980757  
KEYWORDS  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Caprales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 122322)  
Bevan, M., Benes, V., Rechmann, S., Borkova, D., Ansoorge, W., Honkela, J., Mewes, H.W., Mayer, K. and Schell, C.  
Unpublished  
JOURNAL 2 (bases 1 to 122322)  
AUTHORS EU Arabidopsis sequencing project.  
REFERENCE Direct Submission  
TITLE Submitted (08-APR-1998) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@jrc.ac.uk  
FEATURES  
SOURCE location/qualifiers  
1. 122322  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"

gene  
/chromosome="4"  
199. 921  
/gene="F6118.10"  
complement(199. 347)  
/gene="F6118.10"  
complement(199. 347)  
/gene="F6118.10"  
complement(199. 347, 404. 548, 631. 921)  
/gene="F6118.10"  
/note="contains EST gb: T22575, T22317"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CA18185.1"  
/db\_xref="PID:g1283542"  
/db\_xref="PID:g445153"  
/db\_xref="GI:445153"  
/db\_xref="SPTREMBL:O64383"  
/transl\_table="MAOEAGAVKEKEPNSSAAVAATATATADSVKKNQGFSSR  
LMMGIFVRDDPEKRLQYISREAVLSSMRKRSISMRKLTNLTSSVLFELIANG  
YALITPTEDLDMRSPRFLIPMFLIPAYALAYSSIVSFSKMSLFLICKNGCIYD  
RROQKTEKRAERAKINELKERTNYTTQOLIO"  
complement(348. 403)  
/gene="F6118.10"  
complement(404. 548)  
/gene="F6118.10"  
/number=2  
complement(549. 630)  
/gene="F6118.10"  
/number=2  
complement(631. 921)  
/gene="F6118.10"  
/number=3  
1570. 3411  
/gene="F6118.20"  
/number=1  
1570. 3411  
/gene="F6118.20"  
1570. 3411  
/gene="F6118.20"  
/note="similarity to various predicted proteins, Arabidopsis thaliana"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CA18186.1"  
/db\_xref="PID:g1283543"  
/db\_xref="PID:g2980759"  
/db\_xref="GI:2980759"  
/db\_xref="SPTREMBL:O65543"  
/transl\_table="MSRALSSRLNLEGNKIKGLVSDQFYDEALRLYKIKHSIGNG  
FTALIPSVIKACAFQCEPPLGQNLICIKAKADDDTVVNSLSIMYAKFSKIYAVR  
KTYDEMHRDVTYSCITINSCODGLIEAMKLIKERTFGLIPKSELVASLALCTR  
WSSSKVAVRHFALVLDERMQSVLLSTALVDYILKFDHAAAFHVPQMEVKNES  
WTAMISGCYANONYEMGVDFRAMORENLRPNVTLLSVLPACVELNYSGLYEKIG  
FSRHCCHDERLTAAFMVCRGQVSLRVLFEYSKVDVVMSSMSISGYLETCDG  
SEYRNILNORKEGIEANSVTLLAIYASACTNSLLSFASVSHQIICKGMSHILGN  
ALIDMYAKGCSLSAREVEFELTEKDVSSSWINNVGLGHSSEALFEFKGKIGH  
EYDDAFALILSCNAGLVEAQITFTQKIKHMYVTYEHYACTINLGRGKIDDA  
FEYTNMPKPSRIWSSLSACETGRDVAKIIANELMKSPEPNPANYVLKSIH  
TESGNYHAEVEVRVMQRRLNKCYSKIEPELQIEDYOGKMSPI"  
4648. 6966  
/gene="F6118.30"  
complement(4648. 5128, 6668. 6966)  
/gene="F6118.30"  
/note="similarity to TINY, Arabidopsis thaliana, PAVCHX:E218696"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CA18187.1"  
/db\_xref="PID:g1283544"  
/db\_xref="PID:g2980760"  
/db\_xref="GI:2980760"  
/db\_xref="SPTREMBL:O65544"  
/transl\_table="WYIGNOKVAAIGINVSQWITYHGLANVTDTLTFPMNIVPGCI



or send email to webmaster@www-alis.tokyo.jst.go.jp'.

## FEATURES

SOURCE

Location/Qualifiers  
1. .149298  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/clone="T1040-119B8f\_2"  
/map="21q22.1"

BASE COUNT 43483 a 31669 c 30866 g 43280 t  
ORIGIN

Query Match 9.4%: Score 18; DB 9; Length 149298;  
Best Local Similarity 100.0%: Pred. No. 3.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 ACTTGGAATACAGCTCG 84  
Db 36797 ACTTGGAATACAGCTCG 36814

RESULT 9  
AP000134 100000 bp DNA PRI 21-MAY-1999  
LOCUS Homo sapiens genomic DNA of 21q22.1 GART and AML f43D11-119B8  
DEFINITION region, segment 9/10, complete sequence.  
ACCESSION AP000134  
NID 94730903  
VERSION AP000134.1 GI:4730903  
KEYWORDS HTG.  
SOURCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (sites)  
Hattori, M., Ishii, K., Toyoda, A., Shiba, T. and Sakaki, Y.  
Homo sapiens 1,011,894bp genomic DNA of 21q22.1 GART and AML region  
Published Only in Database (1999) In press  
2 (bases 1 to 100000)  
Hirakawa, M., Yamaguchi, H., Imai, K. and Shimada, J.  
Direct Submission  
Submitted (15-APR-1999) to the DDBJ/EMBL/Genbank databases. Mka  
Hirakawa, Japan Science and Technology Corporation (JST), Advanced  
Databases Department, 5-3 Yonbancho, Chiyoda-ku, Tokyo 102-0081,  
Japan (E-mail: mika@tokyo.jst.go.jp, URL: http://www-alis.tokyo.jst.go.jp/,  
Tel: 81-3-5214-8491, Fax: 81-3-5214-8470)

COMMENT This sequence is conducted by Kitasato University JST sequencing  
Laboratory as a JST sequencing team.

Principal Investigator: Yoshiyuki Sakaki Ph.D.  
Phone: +81-3-5449-5622, Fax: +81-3-5449-5445,  
sakaki@hgc.ims.u-tokyo.ac.jp  
Sub-Leader: Tadayoshi Shida Ph.D., Masahira Hattori Ph.D. The  
sequence is submitted by Human Genome Sequencing in ALIS project of  
JST.  
Japan Science and Technology Corporation (JST)  
5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/) or send  
email to webmaster@www-alis.tokyo.jst.go.jp.

FEATURES  
SOURCE Location/Qualifiers  
1. .100000

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"

BASE COUNT 27922 a 22591 c 23016 g 26471 t  
ORIGIN

Query Match 9.4%: Score 18; DB 9; Length 100000;  
Best Local Similarity 100.0%: Pred. No. 3.8;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 ACTTGGAATACAGCTCG 84  
Db 99393 ACTTGGAATACAGCTCG 99410

RESULT 10  
AP000212 100000 bp DNA PRI 14-MAY-1999  
LOCUS Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,  
clone f43D11-119B8, segment 10/12, complete sequence.  
DEFINITION  
ACCESSION AP000212  
NID 94827160  
VERSION AP000212.1 GI:4827160  
KEYWORDS HTG.  
SOURCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 100000)  
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,  
Fujiyama, A., Yada, T., Totoki, Y. and Sakaki, Y.  
Homo sapiens 1,109,292bp genomic DNA of 21q22.1 (REGION:  
D21S226-AML CLONE RANGE: f43D11-119B8)  
Published Only in Database (1999) In press  
2 (bases 1 to 100000)  
Fujiyama, A., Yada, T., Totoki, Y. and Sakaki, Y.  
Direct Submission  
Submitted (10-MAY-1999) to the DDBJ/EMBL/Genbank databases.  
Masahira Hattori, The Institute of Physical and Chemical Research  
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1  
Kitasato, Sagamihara, Kanagawa 228-8535, Japan  
(E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-42-778-9923, Fax: 81-42-778-9924)

COMMENT E. coli transposon insertion: The present data does not contain E.  
coli transposon sequences which integrated in the  
original/previous sequences. We determined the boundary between  
the insertion and genomic sequences experimentally, removed the  
insertion sequences, reconstituted the present data. The sequencing  
project is supported by Japan Science Technology Corporation (JST)  
and the Institute of Physical and Chemical Research (RIKEN).

FEATURES  
SOURCE Location/Qualifiers  
1. .100000

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1"

BASE COUNT 27831 a 22543 c 23121 g 26505 t  
ORIGIN

Query Match 9.4%: Score 18; DB 9; Length 100000;  
Best Local Similarity 100.0%: Pred. No. 3.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 ACTTGGAATACAGCTCG 84  
Db 96791 ACTTGGAATACAGCTCG 96808

RESULT 11  
AP000252 67008 bp DNA PRI 15-MAY-1999  
LOCUS Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,  
clone: T650, complete sequence.  
DEFINITION  
ACCESSION AP000252  
NID 94835621  
VERSION AP000252.1 GI:4835621  
KEYWORDS HTG.  
SOURCE Homo sapiens DNA, clone: T650.  
ORGANISM Homo sapiens





```

repeat_region      /rpt_family="MER74B"
11245. .11464
/rpt_family="MER58B"
repeat_region      /rpt_family="MER58B"
13205. .13513
/rpt_family="AlusX"
complement(15208. .15655)
/rpt_family="LTR7"
repeat_region      /rpt_family="MER91A"
17559. .17712
/rpt_family="MER91A"
17721. .17778
/rpt_family="MER91A"
complement(17844. .18227)
/rpt_family="L2"
complement(18257. .18820)
/rpt_family="MER39"
18826. .19245
/rpt_family="MLTIC"
19974. .20148
/rpt_family="L2"
complement(20224. .20712)
/rpt_family="L1P1"
complement(20713. .21010)
/rpt_family="AluY"
complement(21011. .22272)
/rpt_family="L1P1"
22274. .22398
/rpt_family="MER69A"
22449. .22490
/rpt_family="L2"
22491. .22887
/rpt_family="MLT1I"
22888. .23507
/rpt_family="L2"
complement(23537. .23755)
/rpt_family="MIR"
23812. .23835
/rpt_family="AT_rich"
complement(24184. .24491)
/rpt_family="AlusX"
complement(25266. .25337)
/rpt_family="LTR16A1"
complement(25338. .25722)
/rpt_family="MSTR"
complement(25723. .26049)
/rpt_family="LTR16A1"
26080. .26382
/rpt_family="AluJo"
26383. .26479
/rpt_family="GA-rich"
27124. .27207
/rpt_family="MER96"
complement(27280. .27341)
/rpt_family="MLT1H"
27743. .28028
/rpt_family="L2"
28704. .28759
/rpt_family="TCCA)n"
29469. .29514
/rpt_family="TTR)n"
complement(29515. .29794)
/rpt_family="AlusX"
30833. .30896
/rpt_family="TTC)n"
complement(30898. .31206)
/rpt_family="AluJb"
complement(33150. .33341)
/rpt_family="MIR"
33392. .33501
/rpt_family="TTC)n"
complement(33698. .33807)
/rpt_family="MER102"
34311. .34341
/rpt_family="AT_rich"

```

```

repeat_region      complement(36166. .36292)
/rpt_family="MIR"
repeat_region      /rpt_family="AT_rich"
38097. .38117
/rpt_family="FLAM_A"
complement(38233. .38355)
/rpt_family="L2"
38553. .38650
/rpt_family="L2"
complement(40486. .40588)
/rpt_family="MIR"
41688. .41912
/rpt_family="MIR"
complement(42147. .42440)
/rpt_family="AlusG"
complement(42454. .42761)
/rpt_family="AluY"
43932. .44010
/rpt_family="MADE1"
45644. .45664
/rpt_family="AT_rich"
45918. .46208
/rpt_family="AluJo"
46478. .46671
/rpt_family="MER58A"
complement(46698. .46827)

```

```

Query Match      9.4% Score 18; DB 11; Length 176153;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      97 CAACCTTGACCTGACT 114
Db 149508 CAACCTTGACCTGACT 149525

```

```

RESULT 13
LOCUS      CENS00004
DEFINITION Homo sapiens chromosome 14 clone R-643C12, WORKING DRAFT SEQUENCE,
in ordered pieces.
ACCESSION      AL049778
NID      94972122
VERSION      AL049778.2 GI:4972122
KEYWORDS      HTG; HTGS-PHASE2.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE      1 (bases 1 to 205573)
AUTHORS      Direct Submission
TITLE      Submitted (01-JUN-1999) Genoscope - Centre national de sequenceage
JOURNAL      2, rue Gaston Cremieux - BP 191 91006 Evry cedex - FRANCE (E-mail :
secref@genoscope.cns.fr)
COMMENT      On Jun 2, 1999 this sequence version replaced gi:4760255.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1.205573
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="RPCI-11"
/clone="R-643C12"
BASE COUNT      61781 a 43855 c 41677 g 58260 t

```

## ORIGIN

Query Match 9.4%; Score 18; DB 34; Length 205573;  
 Best Local Similarity 100.0%; Pred. No. 3.6;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 135 TGCTACCTGCCTCCAAAT 152  
 ||||||||||||||||  
 Db 41754 TGCTACCTGCCTCCAAAT 41771

## RESULT 14

LOCUS MLCB2548 38916 bp DNA BCT 12-JAN-1999  
 DEFINITION Mycobacterium leprae cosmid B2548.  
 ACCESSION AL023093  
 MIM 93097216  
 VERSION AL023093.1 GI:3097216  
 KEYWORDS 2-amino-4-hydroxy-6-hydroxymethylidihydropteridine  
 pyrophosphokinase; csbp; dihydronopearin aldolase; folb; folc;  
 folK; folP; ftsH; GTP cyclohydrolase I; hpt; hypoxanthine  
 phosphoribosyltransferase; panc; pantoate--beta-alanine ligase;  
 pps; pseudogene; RLEP; small cold-shock protein; sugar dehydratase;  
 topA; topoisomerase I; transferase.  
 Mycobacterium leprae.  
 Mycobacterium leprae.  
 Mycobacterium leprae.  
 Mycobacterium leprae.  
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Mycobacterium.  
 1 (bases 1 to 38916)  
 Badcock,K. and Churcher,C.M.  
 Unpublished  
 2 (bases 1 to 38916)  
 Parkhill,J., Barrell,B.G. and Randsdram,M.A.  
 Direct Submission  
 Submitted (23-APR-1998) Mycobacterium leprae sequencing project,  
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridg  
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.  
 Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne,  
 Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,  
 France Requests for cosmids should be sent to Karin Eigmeier  
 (kei@pasteur.fr)  
 3 (bases 1 to 38916)  
 Eigmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.  
 Use of an ordered cosmid library to deduce the genomic organization  
 of Mycobacterium leprae  
 Mol. Microbiol. 7 (2), 197-206 (1993)  
 93189700

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

## Notes:

The Sanger Centre is funded to complete the sequence of M. leprae  
 by the Helser Program for Research in Leprosy and Tuberculosis of  
 The New York Community Trust.  
 Work in Paris is supported by the Helser Trust, the Association  
 Française Raoul Follereau and the Groupement de Recherches et des  
 Etudes des Genomes (GIR-GRG).  
 Details of M. leprae sequencing at the Sanger Centre are available  
 on the World Wide Web.  
 (URL, <http://www.sanger.ac.uk/Projects/>)  
 CDS are numbered using the following system eg MLCB33.01c. ML (M.  
 leprae), CB33 (cosmid name), .01 (first CDS), c (complementary  
 strand).

The more significant matches with motifs in the PROSITE database  
 are also included but some of these may be fortuitous. The length  
 in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for  
 CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given  
 where these have been used to deduce the initiation codon. All CDS  
 over 100 codons have been analysed. Gene prediction is based on  
 positional base preference in codons especially where there is an  
 increase in the observed/expected third position G + C. CAUTION:

We may not have predicted the correct initiation codon. Where

## FEATURES

## source

## gene

## CDS

possible we choose an initiation codon (atg, gtg, or tgg) which is  
 preceded by an upstream ribosome binding site sequence (optimally  
 5-13bp before the initiation codon). If this cannot be identified  
 we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the  
 sequenced clone. It may be shorter because we only sequence  
 overlapping sections once, or longer, because we arrange for a  
 small overlap between neighbouring submissions.

## Location/Qualifiers

## 1..38916

/organism="Mycobacterium leprae"  
 /db\_xref="taxon:1769"  
 /clone="cosmid B2548"  
 complement(1..575)  
 /gene="panc"  
 complement(<1..575)  
 /gene="panc"

/note="MLCB2548.01c, panc, probable pantoate--beta-alanine  
 ligase, partial CDS, len: 191 aa; highly similar to many  
 e.g. PANC\_ECOLI P31663 (EC 6.3.2.1) (283 aa), fasta  
 scores:opt: 538 z-score: 706.6 E(): 3.1e-32, 52.7%  
 identity in 165 aa overlap"  
 /codon\_start=1  
 /transl\_table=1  
 /product="pantoate--beta-alanine ligase"  
 /protein\_id="CAA18788.1"  
 /db\_xref="PID:e1289522"  
 /db\_xref="PID:g3097217"  
 /db\_xref="GI:3097217"  
 /db\_xref="SPTREMBL:069524"  
 /translation="MTNSRQPAFNPGEINLYSTRDVTNVSALRHGRMYPTMG  
 ALHNGHLALVRAAKRPGSVVSIIFVNPDLFGAEDLDYPTCFDDLLALRAEDVE  
 IFTPTLAAMPFGHRTVTPGALAFLEGPRNHDGVLTVLKLQIVRPDVEE  
 GEKDYQALVIRQVADLVNDVVVGVPTVA"  
 complement(572..1501)  
 /gene="MLCB2548.02c"  
 complement(572..1501)  
 /note="MLCB2548.02c, unknown, len: 309 aa; equivalent to  
 M. tuberculosis TR:006279 (EMBL:Z95557) MTCY07H7B.19 (303  
 aa), fasta scores: opt: 947 z-score: 1886.5 E(): 0, 67.8%  
 identity in 311 aa overlap"  
 /codon\_start=1  
 /transl\_table=1  
 /product="hypothetical protein MLCB2548.02c"  
 /protein\_id="CAA18789.1"  
 /db\_xref="PID:e1289523"  
 /db\_xref="PID:g3097218"  
 /db\_xref="GI:3097218"  
 /db\_xref="SPTREMBL:069525"  
 /translation="MVQFDGSRARLNTAISTGRVGALELRADQVVACSAVSHSR  
 OMVGFPEPTSVASPEVASAEILLALVPCPEAGLMSGVAVSVPPGIVATISV  
 ANGVGLAOLGKDCIGPLAIHPAMFSGSDSLSCOLRDYFGITKDDVGAIAOS  
 LVLMGEPFPCVVYKATLYSVSPHGNHTVYLALEFVRBALGSELGIGVPP  
 AKGEVDDQDLVVERTVGSLARACENTITQROAGITKLVANGDDALGHLVYALM  
 RIGPELQAVYVNLRLKTRQAHAPYDVEALAP"  
 complement(1511..1514)  
 /note="possible RBS upstream of MLCB2548.02c"  
 complement(1602..2900)  
 /gene="MLCB2548.03c"  
 complement(1602..2900)  
 /note="MLCB2548.03c, possible membrane protein, len: 432  
 aa; equivalent to, but shorter than, M. tuberculosis  
 TR:006278 (EMBL:Z95557) MTCY07H7B.18 (462 aa), fasta  
 scores: opt: 869 z-score: 1209.0 E(): 0, 59.7% identity in  
 432 aa overlap"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative membrane protein"  
 /protein\_id="CAA18790.1"  
 /db\_xref="PID:e1289524"  
 /db\_xref="PID:g3097219"

```

/db_xref="GI:3097219"
/db_xref="SPIREMBL:069526"
/translation="MTVLSRGRVRRGRGRPWLLTVSLLANGASSALVETNRVEL
LKLAVIILAMAAVAFAVSFVRRSDADQSRVNLKLYDLQDRISAREVELYV
ESQRLRELASELRSQADEVALRAELALSTLEFDLQHRPALAGAEAPQAKA
APPCAYSNVRDQSPVDVNSNRVTSDQVTSADSTSIIDVEEPLPRGQ
APQCGVLPHEPDPSSDSREPRHRPRAPIQELLQPMKMDADDEQLPPG
TQDSAMTDESASASVAFSGRRKRSHSIDEAASSPGAEATQOQSGRRANSR
HSAEYRAVSIETGIAKSTPISTOPTAPGAPPELVAPSPSDVPRHRADLLT
DGVGRDLAAGQGVADLMARLGAESTDGRRRRREHYA"
complement(3046..3522)
/gene="MBCB2548.04c"
complement(3046..3522)
/gene="MBCB2548.04c"
/translation="MBCB2548.04c, unknown, len: 158 aa; equivalent to
M. tuberculosis TR:006277 (EMBL:Z95557) MTCY07H7B.17 (158
aa), fasta scores: opt: 944 z-score: 1588.8 E(): 0, 85.48
identity in 158 aa overlap"
/codon_start=1
/translation="hypothetical protein MBCB2548.04c"
/protein_id="CAA18791.1"
/db_xref="PID:e1289525"
/db_xref="PID:g3097220"
/db_xref="GI:3097220"
/db_xref="SPIREMBL:069527"
/translation="MGPTRRKRDLTAAAGAAVGYLLVLYRWPEPTIWTGSLLA
VAIPEALMARVTRKISDGEIDGPGMLHPVAHSIMVAKASAMGALVGMVGYL
VYFLPMPMLRVADKDTSGTVVALLVALVALMHCKSPQDPTHEGAEEN"
complement(3352..4097)
/gene="foik"
complement(3522..4097)
/gene="foik"
/translation="MBCB2548.05c, foik, probable
2-amino-4-hydroxy-6-hydroxymethylidihydropteridine
pyrophosphokinase, len: 191 aa; similar to many e.g.
HPRK_BACSU P28252 bacillus subtilis (167 aa), fasta
scores: opt: 177 z-score: 355.0 E(): 1.2e-12, 33.78
identity in 169 aa overlap"
/codon_start=1
/translation="2-amino-4-hydroxy-6-hydroxymethyl
dihydropteridine pyrophosphokinase"
/protein_id="CAA18792.1"
/db_xref="PID:e1289526"
/db_xref="PID:g3097221"
/db_xref="GI:3097221"
/db_xref="SPIREMBL:069528"
/translation="MTRVLSIGSNLGDRLAMLOSAYDGLGDAVAVSPYDVPKGA
VEKRSFLNAVYIADGPAIDTRKALICRAQELERNRGRVGGRMGAKTIDVDLISCTQS
GATGAVEVITCESNLTLPRAHLRAFLVLPWLVADSDAELTVAGRAQRDRLLAM
EPTRECVRLTNLTILKLRSSPARVPKSD"
complement(4097..4495)
/gene="foib"
complement(4097..4495)
/gene="foib"
/translation="MBCB2548.06c, foib, probable dihydroonepterin
aldolase, len: 132 aa; similar to many e.g. FOLB_BACSU
P28823 dihydroonepterin aldolase (ec 4.1.2.25) (120 aa),
fasta scores: opt: 227 z-score: 337.2 E(): 1.2e-11, 33.98
identity in 118 aa overlap"
/codon_start=1
/translation="dihydroonepterin aldolase"
/protein_id="CAA18793.1"
/db_xref="PID:e1289527"
/db_xref="PID:g3097222"
/db_xref="GI:3097222"

```

Query Match 8.9%; Score 17; DB 1; Length 38916;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 6 TGTAGTGACAAAACA 22
Db 37047 TGTAGTGACAAAACA 37063

RESULT 15
AF058919
LOCUS AF058919 91149 bp DNA PLN 14-APR-1998
DEFINITION Arabidopsis thaliana BAC F6N23.
ACCESSION AF058919
MID G3047100
VERSION AF058919.1 GI:3047100
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (bases 1 to 91149)
Washington University Genome Sequencing Center.
The A. thaliana Genome Sequencing Project
Unpublished (1997)
REFERENCE 2 (bases 1 to 91149)
Getzel, C.
The sequence of A. thaliana F6N23
Unpublished (1998)
3 (bases 1 to 91149)
Waterson, R.
Direct Submission
Submitted (10-APR-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: twilson@watson.wustl.edu

COMMENT
MAPPING: Clones were assigned to the YAC map by hybridization by
M. Lohli, Cold Spring Harbor Laboratories, and fingerprinted
by M. Marra, Washu, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The clone sequenced to the left is F5110. The actual start of this
clone is at base position 104960 of F5110; actual end is at base
position 91149 of F6N23.

NOTES:
Coding sequences below are predicted from computer analysis, using
the program GeneFinder(P. Green and L. Hillier, ms in preparation).
Location/Qualifiers
1..91149
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/clone="F6N23"

```

FEATURES  
 source

```
/chromosome="1Y"
/map="unknown"
/complement(145..907)
/gene="F6N23.23"
/complement(join(145..347,396..512,598..907))
/gene="F6N23.23"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC13624.1"
/db_xref="PID:93047113"
/db_xref="GI:3047113"
/translation="MENLSPNDENDHOSPRSDONDOASVETPYSTMSIDSFYPR
TCSTSGSPDIDETNSFCSEASPCMCVLTESKCIISGEMOSECYAOYISE
PELETMRKRAKLIGEDMSSGKVCYCAVITISNITLNLNCKANDISFDNVCISIVM
TESEVRAVRNNEISIVEERNEISFCMLRHCRIKRLRLYHOODSTLLQI"
/complement(4919..7497)
/gene="F6N23.22"
/complement(join(4919..4993,5078..5488,5567..6055,
6157..6213,6498..6594,6935..7222,7349..7497))
/gene="F6N23.22"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC13623.1"
/db_xref="PID:93047112"
/db_xref="GI:3047112"
/translation="MSLTMDGVEAARSKRNSLKKQLALAVRSQMSYATIFWS
SLTPGVLEWEGCYNGDMKRRKSYESHYRGLOKSKELRLYLSMLEGSGTIVST
THDNLDDDDHSHSMNLSPDLSDEMYLVSMSYVSPQCLPGRASATGETIML
CNAOYAEKRLFSRLARTVVCFLGVIEGLVETLSEDNHLNRKISCLMETSIAH
ONDDEKMEIKISEKHQLPLIGIDEDLHYKRTISTVLANSADSGKNRIRHROP
NIVTSESSFLRKCEKQOYSGFYQKKSQNVLEKILHDPIMTKRMFSONGLN
ODPSDRKNEKESVLTIRVPTVNEVKESLNNITLYLOELKAREVELSEKSVN
FVEORKTENLNDLVLETSNGVDSTKIDNSGETEQVTFEDKTHLRVYKETE
VIVKRCYRDYIVADIMETLSNLMDAFVSRSHTLNKFLTLNLAKFRGAAYASVGM
IKRELRYIGLE"
/10235..12675
/gene="F6N23.1"
/join(10235..10495,10693..10896,10998..11192,11341..11505,
11633..11893,11998..12216,12315..12534,12617..12675)
/gene="F6N23.1"
/notes="similar to the family of glycosyl hydrolases"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC13634.1"
/db_xref="PID:93047123"
/db_xref="GI:3047123"
/translation="MAIRLNSVLPVSKLGAAPTFRASRSLPFSVGMRGVSPFSGA
RPIVLAAYKLAESTEDRVPIDDDDDSTDOVSEFHLDDSLGVIDMSEVPEE
LLDOLRLKSYNVDGVMYDCWKGITESTROYNNSGYKRLQOMIREGLKIQVMSF
HECGNVGDVHIOIPENVREIGQSNPDIYFTDSAGRNTECLTWIDOKRYLGRITA
LEVYFDVMSRFEDEFEEKIIPEIEVGLPCGELRYPSYPAQFGKRYGIGFQIC
YKYLMSNLKAAEVRGHSFWRGPDNTEYNSPHGTFPDGDDYDYSYGRFLNM
YSRYLIDGDRVLAMANLAFEGTCAALSGIHHMYKASHAAELTAGSYNSNDVY
GPVAMEKKHDAALNFCVELRTIJDHEDPEPALDPRIAMOVYNAAMDASIPASE
NALCYDEGYNKILNENKPLTPDPGRHLSCFTYRLNPTLMESNFEFERFLKRMH
GEAVPDLGLAPGTOETNPE"
/complement(13027..14791)
/gene="F6N23.21"
/complement(join(13027..13335,13480..13615,13712..13840,
13943..14001,14087..14791))
/gene="F6N23.21"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC13622.1"
/db_xref="PID:93047111"
/db_xref="GI:3047111"
/translation="MSILCYVPLECYCIGCTHMLKCKLXASGHESENNGLATSDE
FEPIPRICRLILAYEENLHDPMAAPPGYIGDPEHVLKLDYDTERVPIYITLD
HENDGVLAIRGLNLAKEDYAVLNDNLGQTKPDGFGVYHNGLLKAAWVSEEEHVL
RELLEANSYSILTEVGSIGAGVSVLVLVYONRNGVNIERKIRCFALAPPCMS
LHLAVTVAVINSVLODDPLPRTTALLENVRSIICPLCLICLCLDFTFEERKL
KDARIVAPGLRHYHVRKPLRIGRPVPTVAVPDGAPDIOIVSCNATADHAIWI
ERESORALDKIVRKSIVEDHDEETRAAIMKRAASLNINMSPSISGITHDIEGESSA
```

```
GSMEGSPSGWSEFKGMRKMQFDIDCHFPVNDSEHMTFKNOEQALLCGOET"
15539..16687
/gene="F6N23.2"
/join(15539..15554,15672..15764,15953..16026,16133..16208,
16307..16353,16466..16687)
/gene="F6N23.2"
/notes="Similar to protein kinase; coded for by A. thaliana
cDNA H36947; coded for by A. thaliana cDNA H37158"
/codon_start=1
/protein_id="AAC13612.1"
/db_xref="PID:93047101"
/db_xref="GI:3047101"
/translation="MNTSEVYEFIGNVTLIORTSSSLKRISEVYVERGKRGDYVR
ENQVADYVFILOQVRAOVYRSAGEHYDEPLKRRIDFSGHGFEGDYASDVAYTE
LTCLLMSDHRALLLEIKSVSDSKRECLVEDILYLEPLDVAVSPNCSFSSISVLMK
VLTHFYHHSLESEL"
16976..19151
/gene="F6N23.3"
/join(16976..17111,17262..17415,17785..17889,18316..18417,
18618..18666,19056..19151)
/gene="F6N23.3"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC13635.1"
/db_xref="PID:93047124"
/db_xref="GI:3047124"
/translation="MEIYSNCDITVYASQALAAASKVETMTKIVHNEFCYELLVGD
NIPITVDNRIDDNFNFRSVDRKOKKITFTFASPOVSCYIPELTFHYSLDYRNO
VATISVDFPDIFFCEPNRSTKONKSPRCVAFASDLIRATISLNPBREGMSVVA
LSLHSMWFHFRVPRADWLLEVIYSPATSESRGATGMFRKSGEVCIEE"
/complement(20843..21366)
/gene="F6N23.20"
/complement(join(20843..20944,21002..21056,21103..21155,
21292..21366))
/gene="F6N23.20"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC13621.1"
/db_xref="PID:93047110"
/db_xref="GI:3047110"
/translation="MVATAVASFYLTGIDYGPBNVLDPYLSSMILLVEVYIRMF
SYWMLHLEMLVEHMTKRIILAAQDSVKQFIQGPGEPSGKPEPDAK"
/complement(22410..24753)
/gene="F6N23.19"
/complement(join(22410..22521,23220..23538,23620..23939,
24009..24137,24240..24352,24439..24548,24639..24753))
/gene="F6N23.19"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC13637.1"
/db_xref="PID:93047126"
/db_xref="GI:3047126"
/translation="MSPCPSYINVELKPFPHLMODQIITNVYKRYGRNKRIRAECLP
SEENRNDVOCOHMLKLVLDPSLQKGAKEDELSLVEYDYMENDEPSPKSKIEL
PGRIGKCRERHNNHNLNPTIILKSPVREBELIIOAORGNGNKAELAKILPGTEEN
IKRNNCSVYKRLIQPFPSNLFSGVYSGKPSGCEYVFNFRONTMVECSITQIKKAA
KSPQRDLDTLGLNWSISSSTSLGKEEVSVSVCVCRNLACLETPONSNDIV
CVKEVERKERTLRMAARFDIPSIISSTSPASGLKRIKOKYDDPFPPTDARSHMSSE
DHSVASPSSKRYRVRKNTCSGSKPLERBLDFLMDHEGRNGIYVSVRIIPKS
DLKSGIVRPFWR"
/complement(25967..27832)
```

```
Query Match 8.9%; Score 17; DB 8; Length 91149;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 52 ATTGCTAAGTTGGCAC 68
DB 27395 ATTGCTAAGTTGGCAC 27411
```

Search completed: August 6, 1999, 11:39:56

Job time: 6482 sec

This Page Blank (uspto)

---



```

FT      /*tag- a
PN      /note- "no stop codon given"
PD      US5733748-A.
PE      31-MAR-1998.
PF      06-JUN-1995; 469667.
PR      06-JUN-1995; US-469667.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Rosen C, Yu G;
DR      WPI; 98-229823/20.
DQ      P-PSDB; W46879.
PT      Colon-specific nucleic acids - useful as probes for detecting colon
PT      cancer micrometastases
PS      Claim 15; Fig 5A-B; S1pp; English.
CC      V1668-81 represent polynucleotide sequences of partial or full-length
CC      cDNA clones of colon-specific genes. The polynucleotides can be used
CC      as probes to detect expression of the corresponding human genes, e.g. in
CC      diagnostic assays for detecting micrometastases of colon cancer.
CC      Recombinant cells containing the polynucleotides can be used to
CC      produce the polypeptides, in order that antibodies can be raised and
CC      used in further screening or diagnostics.
SQ      Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match          91.6%; Score 141; DB 1; Length 878;
Best Local Similarity 100.0%; Pred. No. 9.3e-65;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACAGATCTTTTCATGCTGATTCAGGCTGTTGATAGGTCGATCGAATCAGAAATATCC 60
DB      545 ACAGATCTTTTCATGCTGATTCAGGCTGTTGATAGGTCGATCGAATCAGAAATATCC 604
QY      61 AACATTGACAGATGATCTTTTATTCCTCCACAGATCCGCCAGAGACACTAGTCT 120
DB      605 AACATTGACAGATGATCTTTTATTCCTCCACAGATCCGCCAGAGACACTAGTCT 664
QY      121 GATGAACGTCGCTCTGT 141
DB      665 GATGAACGTCGCTCTGT 685

RESULT 3
T22483
ID      T22483 standard: cDNA to mRNA; 255 BP.
AC      T22483;
DE      22-AUG-1996 (first entry)
KW      Human gene signature HUMGS04094.
KW      Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW      human; cloning; mapping; non-biased library; diagnosis; detection;
KW      cell typing; abnormal cell function; ss.
OS      Homo sapiens.
PN      W09514772-A1.
PD      01-JUN-1995.
PF      11-NOV-1994; J01916.
PR      12-NOV-1993; JP-355504.
PA      (MATS/) MATSUBARA K.
PA      (OKUBA/) OKUBO K.
PI      Matsubara K, Okubo K;
DR      WPI; 95-206931/27.
PT      Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT      for diagnosis of abnormal cell function, by preparing cDNA that
PT      reflects relative abundance of corresp. mRNA in specific human
PT      tissues
PS      Claim 1; Page 1138-1139; 2245pp; Japanese.
CC      A single-stranded DNA (or its complementary strand or the corresp.
CC      double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC      given in T19001-T26837 and which is able to hybridise to part of
CC      human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC      sequences were obtained from 3'-directed cDNA libraries prepared
CC      from various human tissues; synthesis of cDNA was initiated from the
CC      3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC      untranslated sequence is unique to a particular mRNA species, almost
CC      all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC      is constructed sp as to reflect accurately the relative abundance of

```

```

CC      different mRNAs in the particular tissue from which it was derived.
CC      The appearance frequency of a given GS in a cDNA library can be
CC      determined (esp. using primers and probes derived from the GS
CC      sequences) as a means of diagnosing abnormal cell function or for
CC      recognising different cell types.
SQ      Sequence 255 BP; 76 A; 52 C; 39 G; 80 T;

Query Match          43.5%; Score 67; DB 1; Length 255;
Best Local Similarity 100.0%; Pred. No. 3.8e-26;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      88 CCTCCAGAGCTCCGCCAGAGACACCTAGTCTGATTAAGCTGCTCTGCTCTAT 147
DB      49 CCTCCAGAGCTCCGCCAGAGACACCTAGTCTGATTAAGCTGCTCTGCTCTAT 108
QY      148 ATTGATA 154
DB      109 ATTGATA 115

RESULT 4
Q05590
ID      Q05590 standard: cDNA; 5095 BP.
AC      Q05590;
DE      14-DEC-1990 (first entry)
DE      Insert sequence encoding signal peptide and first 1643 AAs of
DE      mature apob protein.
KW      Atherosclerosis; lipid binding peptide; LBP; phospholipid;
KW      lecithin cholesterol acyl transferase; L-CAT; apo AI;
KW      Chinese hamster ovaries.
OS      Homo sapiens.
PN      Key
PD      Location/Qualifiers
FT      cds      85..5095
FT      signal_peptide 85..165
FT      /*tag- a
FT      /*tag- b

QY      US4943527-A.
DB      24-JUL-1990.
PF      27-FEB-1986; 834300.
PR      04-OCT-1985; US-784418.
PR      04-DEC-1985; US-804692.
PR      27-FEB-1986; US-834300.
PA      (CALB-) CALIFORNIA BIOTECHN.
PI      Procter AA, Vigne JL, Mallory JB, Talmadge KD, Kane JP;
DR      WPI; 90-246622/32.
PT      Prod. of purified lipid-binding peptide - capable of binding to
PT      phospholipid(s), useful in e.g. atherosclerosis.
PS      Disclosure; pp; English.
CC      Lipid-binding peptides in conjunction with phospholipids are
CC      effective in reversing cholesterol transport, and may enhance
CC      lecithin cholesterol acetyl transferase, lowering the rate of
CC      plaque formation and atherosclerosis.
SQ      Sequence 5095 BP; 1430 A; 1281 C; 1179 G; 1205 T;

Query Match          10.4%; Score 16; DB 1; Length 5095;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      89 CTCACAGACTCCGCC 104
DB      3042 CTCACAGACTCCGCC 3057

RESULT 5
N70996
ID      N70996 standard: cDNA; 8247 BP.
AC      N70996;
DE      17-APR-1991 (first entry)
DE      Sequence encoding N-terminal apolipoprotein B (apob).
KW      Lipid-binding peptide; apolipoprotein purification; ss.

```



```
OS Homo sapiens. Location/Qualifiers
FH Key 87.167
FH signal_peptide /*tag-a
FH mat_peptide 168.8247
FH /*tag-b
PN W08702062-A.
PD 09-APR-1987.
PE 02-OCT-1986: U02075.
PR 04-OCT-1985: US-784418.
PR 04-DEC-1985: US-804692.
PR 27-FEB-1986: US-834300.
PA (BIOT-) BIOTECHN RES PARTN.
PI Procter AA, Vigne JL, Mallory JB, Talmadge KD;
DR WPI: 87-108703/15.
DR P-FSDB: P70647.
PT Purified lipid-binding peptide prodn. - by expression of
PT recombinant system and adding lipid to obtd. peptide mixt. to
PS form lipo-peptide complex
PS Disclosure: p73-85; 123pp; English.
CC Apob is a lipid-binding peptide adapted to bind to phospholipids at
CC one or more amphipathic alpha-helical peptide regions. Apob gene is
CC pref. expressed in Chinese hamster ovary (CHO) cells and the gene is
CC introduced in an expression vector having a regulatable promoter
CC derived from the human metallothionein II gene.
SQ Sequence 8247 BP; 2497 A; 1933 C; 1768 G; 2045 T;

Query Match 10.4%; Score 16; DB 1; Length 8247;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 CTCACAGACTCCGCC 104
DB 3042 CTCACAGACTCCGCC 3057
|||||
RESULT 6
VS2153/c
ID V52153 standard; DNA; 21338 BP.
AC V52153:
DE 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:20.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN W09818931-A2.
PD 07-MAY-1998.
PE 30-OCT-1997: U19588.
PE 31-OCT-1996: US-029960.
PA (HWA-) HUMAN GENOME SCI INC.
PA Barsch SA, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI: 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1: Page 257-269; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (VS2153 to VS2524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (VS2153 to VS2524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
```

```
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 21338 BP; 7339 A; 3446 C; 4112 G; 6441 T;

Query Match 10.4%; Score 16; DB 1; Length 21338;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TTCAGGCTGTGATTA 35
DB 5339 TTCAGGCTGTGATTA 5324
|||||
RESULT 7
V69289/c
ID V69289 standard; cDNA; 2832 BP.
AC V69289:
DE 01-MAR-1999 (first entry)
DE Murine JNK-interacting protein 1 (JIP-1) cDNA.
DE JIP-1; JNK-interacting protein; c-Jun NH2-terminal kinase;
KW signal transduction; inhibitor; mouse; neurodegenerative disease;
KW Parkinson's disease; Alzheimer's disease; blood clot; stroke;
KW malignancy; cancer; leukaemia; autoimmune disease; inflammation;
KW apoptosis; therapy; diagnosis; ds.
OS Mus sp.
FH Key
FH Location/Qualifiers
FH CDS 180..2162
FH /*tag-a
PN W09849188-A1.
PD 05-NOV-1998.
PE 28-APR-1998: U08513.
PE 28-APR-1997: US-819177.
PA (UYMA-) UNITV MASSACHUSETTS.
PA Davis RJ, Dickens M;
DR WPI: 99-024042/02.
DR P-FSDB: W81525.
PT c-Jun NH2-terminal kinase (JNK)-interacting protein 1 - used to
PT treat neurodegenerative disease, blood clot, leukaemia, autoimmune
PT disease, and inflammation
PS Claim 1: Page 66-70; 95pp; English.
CC This cDNA sequence codes for murine JNK-interacting protein 1
CC (JIP-1, see W81625), a novel cytoplasmic anchor protein that
CC specifically binds to and inhibits the biological effects of JNK
CC (c-Jun NH2-terminal kinase), including the initiation of apoptosis
CC and oncogenic transformation. To isolate the clone, a yeast
CC two-hybrid method was used to screen a mouse embryo cDNA library to
CC identify proteins that interact with JNK. 7 overlapping partial
CC clones were isolated, and the full-length cDNA was subsequently
CC obtained by screening a mouse brain lambda ZAP1 cDNA library. The
CC invention provides JIP-1 nucleic acids and polypeptides, expression
CC vectors and host cells. The JIP-1 polypeptide and nucleic acids
CC (including antisense and ribozyme molecules) can be used in the
CC manufacture of a medicament for treating a pathological condition
CC associated with abnormal expression or activity of JNK, such as a
CC neurodegenerative disease (selected from Parkinson's disease and
CC Alzheimer's disease), a blood clot, stroke, malignancy, leukaemia,
CC an autoimmune disease or inflammation (all claimed).
SQ Sequence 2832 BP; 569 A; 888 C; 840 G; 535 T;

Query Match 10.4%; Score 16; DB 1; Length 2832;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 CCACAGACTCCGCCGAG 106
|||||
```

Db 2088 CCACAGACTCCGCCAG 2073

RESULT 8

ID X13432/C  
AC X13432:  
DE 19-MAR-1999 (first entry)  
KW Enterococcus faecalis genome contig SEQ ID NO:495.  
KW Enterococcus faecalis contig: detection: Enterococcal infection;  
KW vaccine: attenuation: computer readable medium; ds.  
OS Enterococcus faecalis.  
PN WO9850553-A2.  
PD 12-NOV-1998.  
PF 04-MAY-1998; U08985.  
PR 14-NOV-1997; US-066009.  
PR 06-MAY-1997; US-044031.  
PR 16-MAY-1997; US-046655.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Dillon PJ, Kunsch CA;  
DR WPI: 99-045171/04.  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
PS Claim 1; Page 1739-1742; 2084pp; English.  
CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
SQ Sequence 6645 BP; 1969 A; 1250 C; 1594 G; 1830 T;

Query Match 10.4%; Score 16; DB 1; Length 6645;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 AATCAGAAATATCCA 61

Db 2771 AATCAGAAATATCCA 2756

RESULT 9

ID X13104/C  
AC X13104:  
DE 19-MAR-1999 (first entry)  
KW Enterococcus faecalis genome contig SEQ ID NO:167.  
KW Enterococcus faecalis contig: detection: Enterococcal infection;  
KW vaccine: attenuation: computer readable medium; ds.  
OS Enterococcus faecalis.  
PN WO9850555-A2.  
PD 12-NOV-1998.  
PF 04-MAY-1998; U08985.  
PR 14-NOV-1997; US-066009.  
PR 06-MAY-1997; US-044031.  
PR 16-MAY-1997; US-046655.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Dillon PJ, Kunsch CA;  
DR WPI: 99-045171/04.  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.  
PS Claim 1; Page 939-949; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC X12938 to X13919 represent these nucleotide sequences which are primary  
CC nucleotide sequences, also known as contigs. The computer-based system  
CC can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
SQ Sequence 19031 BP; 6046 A; 3825 C; 4068 G; 5072 T;

Query Match 10.4%; Score 16; DB 1; Length 19031;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 TCTTTTCATTCCTATT 21

Db 14967 TCTTTTCATTCCTATT 14952

RESULT 10

ID N80616  
AC N80616:  
DE 21-NOV-1990 (first entry)  
KW Genomic clone encoding SAP(Phe).  
KW SAP(Val); SAP(Phe); hyaline membrane disease; ss.  
FW Key Location/Qualifiers  
FT exon 1054..1120  
FT /\*tag- a  
FT /label-exon I  
FT 1431..1558  
FT /\*tag- b  
FT /label-exon II  
FT 2060..2131  
FT /\*tag- c  
FT /label-exon III  
FT 2495..2620  
FT /\*tag- d  
FT /label-exon IV  
FT 3447..3635  
FT /\*tag- e  
FT /label-exon V  
FT 3863..3952  
FT /\*tag- f  
FT /label-exon VI  
FT 5386..5568  
FT /\*tag- g  
FT /label-exon VII  
FT 5742..5887  
FT /\*tag- h  
FT /label-exon VIII  
FT 7122..7202  
FT /\*tag- i  
FT /label-exon IX  
FT 7697..7756  
FT /\*tag- j  
FT /label-exon X  
PN WO8804324-A.  
PD 16-JUN-1988.  
PF 03-DEC-1987; U03180.  
PR 03-DEC-1987; WO-U03180.  
PR 03-DEC-1987; US-939206.  
PA (ABBO) Abbott Laboratories.  
PI Whitsett JA, Fox JL, Pilot-Mattias TV, Meuth JL;





Sequence 2783 BP; 565 A; 897 C; 797 G; 520 T;

Query Match 9.7%; Score 15; DB 1; Length 2783;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 TCTGCTCTGTGCT 144  
 |||||  
 DB 835 TCTGCTCTGTGCT 821

Search completed: August 5, 1999, 17:41:12  
 Job time: 6307 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:42:09; Search time 3268.17 Seconds

(without alignments)  
149,860 Million cell updates/sec

Title: US-09-049-696-15

Sequence: 1 ACAGATCTTTTCATGCTAT.....TCTTGCTCTAATTCATA 154

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBank: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_st: \*  
14: gb\_sy: \*  
15: gb\_un: \*  
16: gb\_vl: \*  
17: em\_fun: \*  
18: em\_hcg: \*  
19: em\_hum1: \*  
20: em\_hum2: \*  
21: em\_hum3: \*  
22: em\_hum4: \*  
23: em\_om: \*  
24: em\_or: \*  
25: em\_ov: \*  
26: em\_pat: \*  
27: em\_ph: \*  
28: em\_pl: \*  
29: em\_ro: \*  
30: em\_sts: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_vl: \*  
34: gb\_hcg1: \*  
35: gb\_hcg2: \*  
36: gb\_in1: \*  
37: gb\_in2: \*  
38: em\_ba1: \*  
39: em\_ba2: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result 8  
No. Score Match Length DB ID Description  
1 154 100.0 3311 11 AF039400 Homo sapi AF039400 Homo sapi

2	154	100.0	2826	42	AF127036	AF127036 Homo sapi
3	141	91.6	878	5	I95746	I95746 Sequence 8
4	119	77.3	35278	11	AF039401	AF039401 Homo sapi
5	20	13.0	209071	37	AC005285	AC005285 Drosophila
6	19	12.3	3615	4	D83273	D83273 Cyprinus ca
7	12	12.3	89840	8	ATAC004705	ATAC004705 Arabidops
8	18	11.7	4161	4	AB014462	AB014462 Xenopus l
9	18	11.7	163463	35	AC007462	AC007462 Homo sapi
10	18	11.7	10988	37	CELEF57F10	U23175 Caenorhabdi
11	18	11.0	191134	42	AC006065	AC006065 Homo sapi
12	17	11.0	87500	1	AF027868	AF027868 Bacillus
13	17	11.0	207730	1	BSUB0011	299114 Bacillus su
14	17	11.0	60232	2	AE001272	AE001272 Lactococc
15	17	11.0	1059	2	AF037430	AF037430 Bacillus
16	17	11.0	426	2	AF126295	AF126295 Unidentif
17	17	11.0	517	2	MP070515	U70515 Methylocyst
18	17	11.0	2808	4	NVRETINBP	Y09226 N. viridesc
19	17	11.0	107183	8	AC002983	AC002983 Arabidops
20	17	11.0	1468	8	AF002690	AF002690 Podocarpu
21	17	11.0	95959	8	ATAC006217	AC006217 Arabidops
22	17	11.0	113960	8	ATAC006419	AC006419 Arabidops
23	17	11.0	103495	8	ATAC006585	AC006585 Arabidops
24	17	11.0	58690	8	ATAC007288	AC007288 Arabidops
25	17	11.0	3851	8	BFT43783	U43783 Botryotinia
26	17	11.0	123080	8	F10A2	AF117259 Arabidops
27	17	11.0	1424	8	GM060498	U60498 Glycine max
28	17	11.0	200000	9	AB019439	AB019439 Homo sapi
29	17	11.0	135628	9	HS163016	AL031279 Human DNA
30	17	11.0	159344	9	HS422623	AL031003 Human DNA
31	17	11.0	100368	9	HS14K20	AL008731 Human DNA
32	17	11.0	139378	10	HS64K7	AL003168 Human DNA
33	17	11.0	126581	11	AC003957	AC003957 Homo sapi
34	17	11.0	111682	11	AC005162	AC005162 Homo sapi
35	17	11.0	215286	11	HSU91322	U91322 Human chrom
36	17	11.0	43369	12	MMCKX137	AL021127 Mus muscu
37	17	11.0	250	14	G15445	G15445 human STS S
38	17	11.0	56078	34	HSJ29617	AT009617 Homo sapi
39	17	11.0	188358	35	AC006369	AC006369 Homo sapi
40	17	11.0	30636	36	CEC31E10	Z78539 Caenorhabdi
41	17	11.0	36750	36	CELT15B7	AF022985 Caenorhab
42	17	11.0	66899	37	AC004303	AC004303 Drosophila
43	17	11.0	189036	42	AC006043	AC006043 Homo sapi
44	16	10.4	338534	2	EC00983	U14003 Escherichia
45	16	10.4	11369	2	AE000504	AE000504 Escherich

## ALIGNMENTS

RESULT 1  
AF039400  
LOCUS  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,  
complete cds.  
ACCESSION AF039400  
NID 94009457  
VERSION AF039400.1 GI:4009457  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human CLCA1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

## FEATURES

## SOURCE

1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/gene="hCLCA1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPKSSVFILILHLLEGALNSLIQLNNGYEGIVAIIDPNVP  
EDFTLIQIKDMVTOASLYLFEATGKRRYFNVAIILPEWTKRADYVRPLEYKNA  
DIVLVAISTPGRNDEPTTEONGNGEGERHILTPDTAGKLAIEYPOGRAPVEMAH  
LRMGVDEYNDEKFEYLSNGRIQAVRCSAGITGTINVKKCGGSCYTRKCFPMVTGL  
YKGECEFLQSRQTEKASIMFAQHVDSIVECTEONHNKAPKQKONKLNSTWEYL  
RDSEDFKTTPTPTQPPNPTSLQIGRIYCLVDLKSMSATGNRLNRLNQGLFL  
LQVELGSMGVMTFDSAAHVOSELIQINSQSDRLTAKRLPAAAGGTSICGLRSA  
FTVIRKKYPTDGESEIVLLDGEDNTISGCENEVKOSGAIHTVALGPPSAOELIELSK  
MTGGLQIYASPOVONNGILDAFALSSGNVAVSORSIOLSEKGLTQNSOMNGTVIY  
DSTVGKDTLELITWTQPPQILMDPSGKQGGFYVDKNTKAYLIQIPGAKYTKY  
SLQASSQTLTLVTSRASNAATLPITVTSKNTKDSKPSPLVYANIRGASPTLRA  
SVTALIESVNGKTVLLELDNGAGADATKDGVSRYFTYDINGRSVAVRAGLVN  
AARRVYIPQSGALYIPGMINDENIOMNPPEIPNDKDVQKQVCFSTSSGSEFVAS  
DVNAPRIPDLPPGQITDLKAEIHGSLINLTWAPGDYDHGTAKHYIIRISTSLD  
LRKFNESLOVNTTALIPKRNSEVEFLFKPENTFENGTDLFTAIQAVDKVLSKEI  
SNIRVSLFIPQTPPTPSPDTSAPCPNHNINSLIPGHIILKIMMKWIGELQSLIA

## gene

## CDS

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 154; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 8.5e-80;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACAGATCTTTTCATTGCTATTCAGGCTGTGATTAAGTGCATCTGAATCAGAAATATCC 60  
|||||  
DB 2863 ACAGATCTTTTCATTGCTATTCAGGCTGTGATTAAGTGCATCTGAATCAGAAATATCC 2922  
OY 61 AACATTGCACAGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCT 120  
|||||  
DB 2923 AACATTGCACAGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCT 2982  
OY 121 GATGAAGCTGCTGCTCTGCTCTAATATTCATA 154  
|||||  
DB 2983 GATGAAGCTGCTGCTCTGCTCTAATATTCATA 3016

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (caccl1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Agnel, M., Vernat, J., and Culouscou, J.-M.  
AUTHORS Cloning of three human homologs of bovine epithelial chloride  
TITLE

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel, M., and Culouscou, J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Neuilly-Malmaison 92500, France

FEATURES  
SOURCE  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine: colon"  
1. .2826  
/gene="CaCC1"  
5. .2749  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AA025487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"

## gene

## CDS

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 154; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 8.4e-80;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACAGATCTTTTCATTGCTATTCAGGCTGTGATTAAGTGCATCTGAATCAGAAATATCC 60  
|||||  
DB 2516 ACAGATCTTTTCATTGCTATTCAGGCTGTGATTAAGTGCATCTGAATCAGAAATATCC 2575  
OY 61 AACATTGCACAGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCT 120  
|||||  
DB 2576 AACATTGCACAGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCT 2635  
OY 121 GATGAAGCTGCTGCTCTGCTCTAATATTCATA 154  
|||||  
DB 2636 GATGAAGCTGCTGCTCTGCTCTAATATTCATA 2669

RESULT 3  
195746 878 bp DNA PAT 17-JUL-1998  
LOCUS Sequence 8 from patent US 5733748.  
DEFINITION  
ACCESSION 195746  
NID 93940216  
VERSION 195746.1 GI:3940216  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Yu, G. and Rosen, C.  
AUTHORS Unclassified.  
TITLE





BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN

Query Match 77.3%; Score 119; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 2.8e-59;  
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACAGATCTTTTCATTCGATTCAGGCTGTTGATTAAGTCGATCTGAATCAGAAATATCC 60  
DB 33077 ACAGATCTTTTCATTCGATTCAGGCTGTTGATTAAGTCGATCTGAATCAGAAATATCC 33136  
OY 61 AACATTGCACGAGTATCTTTGTTTATCTCTCCACAGACTCCGCCAGAGACACCTAGTCC 119  
DB 33137 AACATTGCACGAGTATCTTTGTTTATCTCTCCACAGACTCCGCCAGAGACACCTAGTCC 33195

RESULT 5  
AC005285 209071 bp DNA INV 15-JUL-1998

LOCUS AC005285 Drosophila melanogaster DNA sequence (pis DS00121 (D128), DS05470 (D270), and DS00108 (D120)), complete sequence.

AC005285 AC004272 AC002637 AC002638 AC003139 AC003717 AC003718  
AC003140 AC003141 AC002639 AC003142 AC003720 AC003143  
AC004568 AC004269 AC002586 AC002587 AC003130 AC002588 AC002589  
AC002590 AC003131 AC003709

NID 93320127  
VERSION AC005285.1 GI:3320127  
KEYWORDS HTG.

ORGANISM Drosophila melanogaster (Subclones in sac from P1 clones DS00121 (D128), DS05470 (D270), and DS00108 (D120)) DNA.

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 209071)  
Celniker, S.E., George, R.A., Galle, R.F., Hoskins, R.A.,  
Svirskas, R.R., Harris, N.L., Agdayani, A., Arcaina, T.T., Baxter, E.,  
Blazer, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E.,  
Flanagan, J., Houston, K.A., Hummasti, S.R., Katta, K., Kearney, L.,  
Kim, S.H., Lee, B., Lomocan, M.A., Mak, J., Mazda, P., Mok, M.S.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclob, J.M., Park, S.,  
Pfeiffer, B., PUNCH, E., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R.,  
Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.

TITLE Direct Submission  
JOURNAL Submitted (15-JUL-1998) Berkeley Drosophila Genome Project, MS  
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, US  
REFERENCE Sequence submitted by:  
AUTHORS Lawrence Berkeley National Genome Project

TITLE Direct Submission  
JOURNAL Submitted (15-JUL-1998) Berkeley Drosophila Genome Project, MS  
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, US  
COMMENT Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

FEATURES  
SOURCE  
1. 209071  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2L"

BASE COUNT 59585 a 45615 c 45383 g 58488 t  
ORIGIN

Query Match 13.0%; Score 20; DB 37; Length 209071;  
Best Local Similarity 100.0%; Pred. No. 0.29;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 62 ACATTGCACGAGTATCTTTG 81  
DB 116525 ACATTGCACGAGTATCTTTG 116544

RESULT 6  
D83273 3615 bp mRNA VRT 06-FEB-1999

LOCUS D83273 Cyprinus carpio mRNA for jnka, complete cds.  
DEFINITION D83273  
ACCESSION D83273  
NID 91434897  
VERSION D83273.1 GI:1434897

KEYWORDS jnka.  
CYPRINUS CARPIO adult ovary cDNA to mRNA.  
ORGANISM Cyprinus carpio

Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Cyprinidae; Cyprinidae; Cyprinus.

REFERENCE 1 (bases 1 to 3615)  
Submitted (01-FEB-1996) to the DDBJ/EMBL/Genbank databases. Hisashi  
Hashimoto, Kyoto University, Faculty of Agriculture, Fishery;  
Otake-cho, Kitashirakawa, Sakyo, Kyoto, Kyoto 606-01, Japan  
(Tel:075-753-6213, Fax:075-753-6223)

REFERENCE 2 (bases 1 to 3615)  
Hashimoto, H., Toyohara, H., Yokoyama, Y., Matsuo, Y. and Sakaguchi, M.  
Molecular cloning of carp mitogen-activated protein kinase (JNK1)  
cDNA and its unique expression in ovary  
Unpublished (1996)

JOURNAL Hashimoto, H., Matsuo, Y., Yokoyama, Y., Toyohara, H. and Sakaguchi, M.  
Structure and expression of carp mitogen-activated protein kinases  
homologous to mammalian JNK/SAPK  
J. Biochem. 122 (2), 381-386 (1997)  
96021077

TITLE Direct Submission  
JOURNAL Location/Qualifiers

FEATURES  
MEDLINE  
JOURNAL  
MEDLINE  
SOURCE  
1. 3615  
/organism="Cyprinus Carpio"  
/db\_xref="taxon:7962"  
/dev\_stage="adult"  
/tissue\_type="ovary"  
1. 208  
209..1492  
/codon\_start=1  
/product="Jnka"  
/protein\_id="BA011880.1"  
/db\_xref="PID:d1012547"  
/db\_xref="PID:g1434898"  
/db\_xref="GI:1434898"

/translation="MNKMKREFYGVNDGDSFTYKRYONRPGSGAGOCIGSAY  
DHLERNVAKRISRPONOTAKKRYRELYVLMKGIKHLHSGIIRHDLKPSNIYK  
DYLVELMDANICQVIOELDERISYLLQKICIKHLHSGIIRHDLKPSNIYK  
SDCTLKILDFGLARATGILMTPVYATRYRKLAVLYENRPRYTGYSPEKLEP  
VRKILPFGEDYIDQNNKYIEQIGPTQEFYMKLVSVRYENRPRYTGYSPEKLEP  
DYVFPADSEHNKAKASOARDLISKMLVIDASKRISVEALOHPIYVWVDPAVEAP  
PVTIDKDLDERHTEVEKELIKYEVLDLDERKNGAINGQSPPLGAAVINGSPQSS  
SSINDVSKSTPTVASDTSLEASAPLSCCR"

3'UTR 1493 .3615  
 polyA\_signal 3597 .3602  
 BASE COUNT 1027 a 672 c 758 g 1158 t  
 ORIGIN

Query Match 12.3% Score 19; DB 4; Length 3615;  
 Best Local Similarity 100.0%; Pred. No. 0.98;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 CGAGTATCTTGTATTATC 88  
 |||||||  
 DB 3183 CGAGTATCTTGTATTATC 3201

RESULT 7  
 LOCUS ATAC004705/c 89840 bp DNA PLN 09-SEP-1998  
 DEFINITION Arabidopsis thaliana chromosome II BAC F26C24 genomic sequence,  
 complete sequence.  
 ACCESSION AC004705  
 MID 93252804  
 VERSION AC004705.1 GI:3252804  
 KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euhylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsids.  
 1 (bases 1 to 89840)  
 Rounsley,S.D., Kaul,S., Lin,X., Ketchum,K.A., Crosby,M.L.,  
 Brandon,R.C., Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D.,  
 Somerville,C.R. and Venter,J.C.  
 Arabidopsis thaliana chromosome II BAC F26C24 genomic sequence  
 Unpublished  
 2 (bases 1 to 89840)  
 Rounsley,S.D. and Lin,X.  
 Direct Submission  
 Submitted (20-MAY-1998) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 On Jun 24, 1998 this sequence version replaced gi:3242697.  
 Address all correspondence to:  
 Steve Rounsley  
 The Institute for Genomic Research  
 9712 Medical Center Dr.,  
 Rockville, MD 20850,  
 USA  
 REFERENCE  
 AUTHORS Rounsley,S.D.  
 TITLE Direct Submission  
 JOURNAL  
 TITLE Submitted (24-JUN-1998) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 COMMENT  
 Address all correspondence to:  
 Steve Rounsley  
 The Institute for Genomic Research  
 9712 Medical Center Dr.,  
 Rockville, MD 20850,  
 USA

e-mail: rounsley@tigr.org  
 BAC clone F26C24 is from Arabidopsis chromosome II and is near the  
 molecular marker m1398.  
 The orientation of the sequence is from SP6 to T7 end of the BAC  
 clone.  
 Genes were identified by a combination of three methods: Gene  
 prediction programs including GRAT (available by anonymous ftp  
 from arthur.epm.ornl.gov), Genefinder (Phil Green, University of  
 Washington), Genscan (Chris Burge/GENSCANW.html), and NetPlantGene  
 (http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the  
 complete sequence against a peptide database and the Arabidopsis  
 EST database at TIGR (http://www.tigr.org/tidb/at.html).  
 Annotated genes are named to indicate the level of evidence for  
 their annotation. Genes with similarity to other proteins are named  
 after the database hits. Genes without significant peptide  
 similarity but with EST similarity are named as 'unknown' proteins.  
 Genes without protein or EST similarity, that are predicted by more  
 than two gene prediction programs over most of their length are  
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are

FEATURES  
 source predicted by tRNAscan-SE (Sean Eddy,  
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are  
 identified by RepeatMasker (Arian Smit,  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of  
 genomic sequence that are not annotated as genes but have predicted  
 exons by GRAT are annotated as misc features.  
 Location/Qualifiers

1..89840  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="II"  
 /map="m1398"  
 /clone="F26C24"  
 15..41  
 /note="exon predicted by xgrail, quality  
 marginal\_shadowexon"  
 78..147  
 /note="exon predicted by xgrail, quality  
 excellent\_shadowexon"  
 complement(123..188)  
 /note="exon predicted by xgrail, quality marginal"  
 complement(201..223)  
 /note="exon predicted by xgrail, quality marginal"  
 337..348  
 /note="exon predicted by xgrail, quality  
 marginal\_shadowexon"  
 complement(<977..>1852)  
 /gene="F26C24.1"  
 complement(<977..>1852)  
 /note="predicted by genscan and genefinder"  
 complement(977..1852)  
 /gene="F26C24.1"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAC24175.1"  
 /db\_xref="PID:93252805"  
 /db\_xref="GI:3252805"  
 /translation="MGSDQCFSRUNSLERKALIKYQKIGHQPADYFPDQGLFSLRIS  
 KSEPDFKSKTVGRENISLHNRILVRSILKNASVAKSPPEPYRKSILGDPVPPSPRK  
 CSRKRFRRPSPGLPGPOSITITNDSMSKARLPPEVSVSDGEVEGDMGSPSY  
 OSRSFUPRPLGVSPHAKSKARFSTYRNGINRPTCCSSGLPMTILRLARLEKLMDEGI  
 KLSDSANLNRGNATYRKLRLPECLSLASQKRAVSNWMLDTHAAENVPRVGEZ  
 WPILEKICQVASEE"  
 complement(1871..1940)  
 /note="exon predicted by xgrail, quality good"  
 complement(2023..2062)  
 /note="exon predicted by xgrail, quality good"  
 complement(2346..2602)  
 /rpt\_family="POLY\_A"  
 2716..2739  
 /rpt\_family="AT\_rich"  
 complement(3450..3538)  
 /rpt\_family="(TAAA)n"  
 3563..3712  
 /note="exon predicted by xgrail, quality  
 excellent\_shadowexon"  
 complement(3870..3893)  
 /rpt\_family="AT\_rich"  
 4276..4450  
 /rpt\_family="AT\_rich"  
 4677..4801  
 /rpt\_family="(TAAA)n"  
 5124..5167  
 /note="exon predicted by xgrail, quality good\_shadowexon"  
 complement(5337..5427)  
 /rpt\_family="AT\_rich"  
 complement(5528..5622)  
 /note="exon predicted by xgrail, quality excellent"  
 5666..5808  
 /note="exon predicted by xgrail, quality  
 excellent\_shadowexon"

```

repeat_region complement(5853.5883)
/rpt_family="AT_rich"
repeat_region 5919.6032
/rpt_family="(TAAA)n"
repeat_region complement(6246.6272)
/rpt_family="AT_rich"
repeat_region 6537.6594
/rpt_family="AT_rich"
repeat_region complement(6779.6857)
/rpt_family="AT_rich"
repeat_region 7213.7230
/note="exon predicted by xgrall, quality good"
misc_feature complement(7556.7704)
/note="exon predicted by xgrall, quality
marginal_shadowexon"
misc_feature 7594.7910
/note="exon predicted by xgrall, quality excellent"
misc_feature complement(7778.7883)
/note="exon predicted by xgrall, quality
excellent_shadowexon"
repeat_region complement(8011.8138)
/rpt_family="(TAAA)n"
repeat_region complement(8088.8197)
/rpt_family="(TAAA)n"
repeat_region complement(8303.8384)
/rpt_family="(TA)n"
repeat_region complement(8429.8488)
/rpt_family="AT_rich"
misc_feature 8531.8563
/note="exon predicted by xgrall, quality marginal"
misc_feature 8648.8826
/note="exon predicted by xgrall, quality good"
repeat_region complement(8979.8999)
/rpt_family="AT_rich"
misc_feature 9036.9125
/note="exon predicted by xgrall, quality marginal"
repeat_region complement(9134.9200)
/rpt_family="AT_rich"
repeat_region complement(9300.9424)
/rpt_family="(TAAA)n"
repeat_region complement(9506.9541)
/rpt_family="AT_rich"
repeat_region complement(10195.10236)
/rpt_family="AT_rich"
repeat_region 10536.10617
/rpt_family="(TAGA)n"
misc_feature complement(10718.10878)
/note="exon predicted by xgrall, quality excellent"
repeat_region 10887.10960
/rpt_family="(TAAA)n"
repeat_region complement(11784.11840)
/rpt_family="AT_rich"
repeat_region complement(12443.12481)
/rpt_family="AT_rich"
repeat_region 12807.12926
/rpt_family="AT_rich"
repeat_region 13019.13073
/rpt_family="AT_rich"
repeat_region 13131.13254
/rpt_family="(TAAA)n"
repeat_region 13437.13464
/rpt_family="AT_rich"
repeat_region complement(13587.13640)
/rpt_family="AT_rich"
repeat_region complement(13649.13777)
/rpt_family="(TAAA)n"
misc_feature 14034.14068
/note="exon predicted by xgrall, quality
marginal_shadowexon"
repeat_region complement(14070.14120)
/rpt_family="AT_rich"

```

Query Match \* 12.3% Score 19; DB 8; Length 89840;

```

Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 77 CTTGTTATCTCTCACA 95
Db 42019 CTTGTTATCTCTCACA 42001

RESULT 8
AB014462/c AB014462 4161 bp mRNA VRT 06-FEB-1999
LOCUS Xenopus laevis XNLR-1 mRNA, complete cds.
DEFINITION AB014462
ACCESSION 63157803
NID AB014462.1 GI:3157803
VERSION neuronal leucine-rich repeat protein; XNLR-1.
KEYWORDS Xenopus laevis cDNA to mRNA.
SOURCE Xenopus laevis
ORGANISM
Enkalyota; Metazoa; Chordata; Vertebrata; Amphibia; Batrachia;
Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus.

REFERENCE
AUTHORS Hayata,T.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-1998) to the DDBJ/EMBL/Genbank databases.
TadaYoshi Hayata, University of Tokyo, Graduate School of Arts and
Sciences; 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
(E-mail:cc87723@komaba.ecc.u-tokyo.ac.jp, Tel:81-3-3468-4080,
Fax:81-3-3468-4087)
2 (sites)
REFERENCE
AUTHORS Hayata,T., Uochi,T. and Asashima,M.
TITLE Molecular cloning of XNLR-1, a Xenopus homolog of mouse neuronal
leucine-rich repeat protein expressed in the developing Xenopus
nervous system
JOURNAL Gene 221 (1), 159-166 (1998)
MEDLINE 99070063
FEATURES
source location/Qualifiers
1..4161
/organism="Xenopus laevis"
/db_xref="taxon:8355"
155..2311
/gene="XNLR-1"
155..2311
/gene="XNLR-1"
/codon_start=1
/product="neuronal leucine-rich repeat protein"
/protein_id="BAA28530.1"
/db_xref="PID:d1029470"
/db_xref="PID:g3157804"
/db_xref="GI:3157804"
/translation="MARISYLVITACQLLGLITSSDSSAPANECPOLCCEIRPM
FTQSTYRENTVDCNDLRTKIPANLSDTQVLLQSNNGICNGELQRLVNLTELD
LSQNFSTSHDVLGSLNLSQTLTLEENQIFETWDCLDNLQELVILNHNQNSIS
PSAFSGLRNLRLHLSNKRIRIDSMPSTPRLITLIMGENVQILDLNFOPLVNL
BSLVLAGMTSDIPGNALIGLNLSELSYDNKVGKVPOLAKVNLNFEIDLNKPI
RKIQEGDFKMLKLEGINNAEIVSDRNAMENPEFKLEATNNPLSTYHRBAF
KNVFTLESILNNALNSYRGVESLPNIREISHSNPLRCDVCVHMGSNOTSIRF
MEPLSMFCALPEYRGQPVKEALADQPAEQCPMTSOTPFSLDILGMTISLDCR
ATAEPEPEIYWTPLGHKVTLETLISDKYHLSGEGSLQIENVVEDSGRTVCVQNSG
ADTKVATLRVNGTLDGTOALRLYVQAESSSVLYKSSVYLAISNKKSSATMKID
NPHTITARVPADVHEYNLHLOPATEYECVLSGLHQAQACINVTTKGYSYSLT
VNDQERSAIAAVMGSLFALLISPAVSVAARPOKRNTRHSLKKYMOTSSIPNLNL
YPLISLWEDSKKEKEGTAESKASVDTSRTYTM"
polya_site
4161
/note="18 A nucleotides"
BASE COUNT 1135 a 948 c 881 g 1197 t
ORIGIN

```

Query Match 11.7%; Score 18; DB 4; Length 4161;  
 Best Local Similarity 100.0%; Pred. No. 3.8;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 132 TGCTCTTGTCTTAAT 149



PHSPHRLADVEERYSGHVEYVNNVRETRASIAHMYIVSEFLIPPLQCI  
 PTVLHGLFJMALTSLSGEMERLLLTEDQAPPTIRVPRKXHLFTACQL  
 LQIILICAFESFYPFIVCFEFLIPRLILPRLIDYKILDALDRH"

BASE COUNT  
 ORIGIN

3422 a 2024 c 1936 g 3606 t

Query Match 11.7%; Score 18; DB 37; Length 10988;  
 Best Local Similarity 100.0%; Pred. No. 3.9;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 ATCTGAATCAGAAATAT 58  
 Db 4799 ATCTGAATCAGAAATAT 4816

RESULT 11  
 AC006065/c  
 LOCUS Homo sapiens 191134 bp DNA PRI 27-FEB-1999  
 DEFINITION Institute Human BAC Library) complete sequence.  
 ACCESSION AC006065  
 NID 94309921  
 VERSION AC006065.3 GI:4309921  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
 AUTHORS

1 (bases 1 to 191134)

Muzny,D., Durbin,J., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,  
 Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,U.H.,  
 Gorrell,U.L., Hernandez,U., Issar,A., Jackson,L., Kneitz,S.,  
 Kondejewski,N., Lau,S., Leal,B., Lee,E., Licharge,O., Liu,W.,  
 Logan,O., Lu,J., Mondel,I., Martinez,C., Merscher,S., Miller,A.,  
 Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,  
 Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shum,C., Simon,M.,  
 Vo,Q., Williamson,A., Worley,K.C., Xiang,A.M., Yang,R., Yu,W.,  
 Zhou,X., Kuchelapatti,R., Nelson,D. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 191134)

Worley,K.C.

Direct Submission

Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 191134)

Worley,K.C.

Direct Submission

Submitted (27-FEB-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On Mar 1, 1999 this sequence version replaced gi:4263064.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
 gc-help@bcm.tmc.edu

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence

continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated error rate less than 1 per 10,000 bases.  
 Reports of lowest quality individual bases and measures of base  
 quality are listed below. Description of the metrics can be found  
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT-----

----- Summary Statistics -----  
 Contig length: 191134  
 Phrap values in estimate: 190243  
 Average error rate (BCM-Phrap estimate): 0.79361e-05  
 Fraction of Phrap values less than 40 : 0.0112658  
 Number of consensus changing edits: 22  
 Number of N's in consensus : 0

Position	Original/Context	Edited/Context
19187	attatataa(t)tcataat	attatataa(a)tcataat
19239	ataataga(g)tatataat	ataataga(a)tatataat
19264	ataataga(g)tatataat	ataataga(a)tatataat
19265	ataataga(g)tatataat	ataataga(a)tatataat
19274	ataataga(g)tatataat	ataataga(a)tatataat
19297	ataataga(g)tatataat	ataataga(a)tatataat
19320	ataataga(g)tatataat	ataataga(a)tatataat
19322	ataataga(g)tatataat	ataataga(a)tatataat
19382	ataataga(g)tatataat	ataataga(a)tatataat
32535	ataataga(g)tatataat	ataataga(a)tatataat
32536	ataataga(g)tatataat	ataataga(a)tatataat
50358	ataataga(g)tatataat	ataataga(a)tatataat
51304	ataataga(g)tatataat	ataataga(a)tatataat
51305	ataataga(g)tatataat	ataataga(a)tatataat
51307	ataataga(g)tatataat	ataataga(a)tatataat
55508	ataataga(g)tatataat	ataataga(a)tatataat
98778	ataataga(g)tatataat	ataataga(a)tatataat
129627	ataataga(g)tatataat	ataataga(a)tatataat
138215	ataataga(g)tatataat	ataataga(a)tatataat
148469	ataataga(g)tatataat	ataataga(a)tatataat
148471	ataataga(g)tatataat	ataataga(a)tatataat

----- Distribution of Quality < 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
10001	9001	*							
9001	8001	*							
8001	7001	*							
7001	6001	*							
6001	5001	*							
5001	4001	*							
4001	3001	*							
3001	2001	*							
2001	1001	*							
1001	0	*	*	*	*	*	*	*	*

Phrap Value Range

```

-----
Version: 1.01 gxf0.
FEATURES
Source
Location/Qualifiers
1. 191134
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPC111-407A16"
/chromosome="12q24.2"
1. 200
/note="Overlaps bases 95981 to 96180 in AC005868."
/function="Overlap with adjacent clone AC005868."
misc-feature
repeat_region
/rpt_family="AluY"
2286. 2472
/rpt_family="MIR"
complement(3519. 3780)
/rpt_family="AluX"
repeat_region
complement(3835. 4004)
/rpt_family="L1M4"
repeat_region
complement(4409. 5123)
/rpt_family="L1M4"
complement(5156. 5283)
/rpt_family="L1P8"
repeat_region
6780. 7077
/rpt_family="AluX"
500. 7800
/rpt_family="AluX"
7842. 8246
/rpt_family="L1M4"
complement(8430. 8527)
/rpt_family="MER85"
9533. 9842
/note="Region: Similar to clone q131a09.x1 Homo sapiens
cDNA, A1224910."
10088. 10405
/note="Region: Similar to clone qd93h03.x1 Homo sapiens
cDNA, A1125490."
10515)
/rpt_family="L1P8"
11258. 11361
/rpt_family="MIR"
complement(12174. 12263)
/rpt_family="MIR"
complement(12427. 12532)
/rpt_family="L2"
complement(12730. 13066)
/rpt_family="AluYb8"
13092. 13188
/rpt_family="MIR"
complement(14949. 15086)
/rpt_family="L1M4"
complement(15046. 15333)
/rpt_family="L1M4"
complement(15518. 15840)
/rpt_family="L1M4"
16533. 16781
/standard_name="D12S2444"
/db_xref="GDB:9785935"
complement(17717. 18009)
/rpt_family="AluUo"
complement(19048. 19436)
/rpt_family="AT-rich"
19140. 19790
/function="Low coverage"
19646. 19802
/rpt_family="AluSg/X"
complement(19805. 19843)
/rpt_family="(7A)n"
complement(19865. 20271)
/rpt_family="L1M4"
complement(20320. 20411)
/rpt_family="L1M4"

```

```

Query Match 11.7%; Score 18; DB 42; Length 191134;
Best Local Similarity 100.0%; Pred.No. 4.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 TCTTTCATGCTATCA 23
|||||
Db 17707 TCTTTCATGCTATCA 17690

RESULT 12
AF027868/c 87500 bp DNA BCT 15-NOV-1997
LOCUS Bacillus subtilis chromosome region between tect and odhA.
DEFINITION AF027868
ACCESSION AF027868
NID 92618993
VERSION AF027868.1 GI:2618993
KEYWORDS
SOURCE
ORGANISM
Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
1 (bases 59208 to 59540)
Muller, J., Walter, F., van Dijk, J.M. and Behnke, D.
Suppression of the growth and export defects of an Escherichia coli
secA(Ts) mutant by a gene cloned from Bacillus subtilis
Mol. Gen. Genet. 235 (1), 89-96 (1992)
93062812
2 (bases 34592 to 35233)
Wolf, M., Geçiz, A., Simon, O. and Boriss, R.
Genes encoding xylan and beta-glucan hydrolysing enzymes in
Bacillus subtilis: characterization, mapping and construction of
strains deficient in lichenase, cellulase and xylanase
Microbiology 141 (Pt 2), 281-290 (1995)
95219081
JOURNAL
MEDLINE
3 (bases 1 to 87500)
Lapidus, A., Galleron, N., Sorokin, A. and Ehrlich, D.
Sequence analysis of the Bacillus subtilis chromosome region
between the tect and odhA loci cloned in a yeast artificial
chromosome
Unpublished
4 (bases 1 to 87500)
Lapidus, A., Galleron, N., Sorokin, A. and Ehrlich, D.
Direct Subcloning
Submitted (01-Oct-1997) Laboratoire de Genetique Microbienne, INRA,
Domaine de Vilvert, Jouy-en-Josas cedex 78352, France
FEATURES
Source
1. 87500
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
<1. 560
/gene="yoxB"
/citation={3}
<1. 560
/gene="yoxB"
/citation={3}
/codon_start=3
/transl_table=11
/product="yoxB"
/protein_id="BAB84418.1"
/db_xref="GI:2618994"
/db_xref="PID:92618994"
/translacion="AETPAASLLGGTLEOMAVDTKTAAGSGSGMRNLTMTDLG
LPOAFGIQSAKLFEGELRHMLHTTSVIRPVFTQKNTGKPAITSPITSTAFGH
KPEALNHTGPAALIPGKAATVCEITIRHSHLONLCLNGFPHSGANGHRKQFS
KKKEOLEROIIRSFALVDFAIKKR"
604. 1212
/gene="yoxA"
604. 1101
/note="yoxA"
/note="Similar to B.subtilis P20 protein and E.coli
ribosomal protein alanine acetyltransferase (194 aa)"
/citation={3}
/codon_start=1

```

```
terminator
gene
CDS
/translation="putative alanine acetyl transferase"
/protein_id="AAB84419.1"
/db_xref="PID:g2618995"
/db_xref="GI:2618995"
/translation="MPEILETDRLLRLKQITPDQDAAFACRSNDEVTRYVLEWSEI
EQATSMIOETPALYQERKGIKRWGIERRDTKELIGTGFHALAQHRRRAIEGTEIPEH
WRNGFASVIVSVYSGFALGSLRIGAVETDNEASNRLLKMGFOKEGVLQRYMTS
KWDSI"
1187. .1212
/gene="yoaA"
/complement(1217. .2461)
/gene="yoaB"
/complement(1217. .2461)
/gene="yoaB"
/translation="similar to B.subtilis multidrug resistance protein
2 (400 aa)"
/codon_start=1
/translation="putative transporter"
/protein_id="AAB84444.1"
/db_xref="PID:g2619020"
/db_xref="GI:2619020"
/translation="MDDKIGIPKRLANGFLGVLFMGDGLGEGMLSPLEIENCLTVO
OSATIFSTIGIALIASWFSVGLAEFAKRTMGLFVIGTAATVGEQNLIP
VMATVYFVKGIGYPLFAYSFVTWYIRTPQSLSTAVGWEMIAVCLMFTVGAMYSY
AIKAFGYLNTLMSISFWVCLAFALPINKDRPKRKRETAELKGVTLILEFTN
RVLGGLIRIINSIGTGFVFLPMMHAGHISYNWMLQIIGTIFLGNINLFTGIV
GDFGKNTVIMFGVGIGITVLLYAPVSGGSLAVSVIGFMGILAGVPIA
IVPVAKDKGAASVNLATAGLSAFVGPALAMFLIGVAGQSVWIFPALYLSAVL
TKCHIEPKVKEETSPQYAS"
complement(2555. .4018)
/gene="yoaC"
/citation(3)
/gene="yoaC"
/translation="similar to Lactobacillus xylose kinase (501 aa)"
/codon_start=1
/translation="xylose kinase"
/protein_id="AAB84445.1"
/db_xref="PID:g2619021"
/db_xref="GI:2619021"
/translation="MKROKGYLVEDIGTGNARNAVSVTGSVLVEREDIEYSTETLY
PDSRYFSPVIMKOVNMLAKRLRSCDIDILGSTSORQIVLIDONGPFLGDN
IDNRGREGAIPMEIEYSTGRPLTALFALSTYGLKOROPSLMERTASFTISNW
VTYOLGILTYEPQATETLLFDYKONTMSEMGDIFGSPSILPIVRAGTAGTGA
NEVASEGLSINAKVIAGGGDTOLAVKSTGLEDIVYSGTTPITKITEBDGDTA
KAWLNCHTDGHWLVETNPGLTGLNKLQIFPNPTYEVEEIEISALKAKEDACVA
ALGSYLSAEKNALRGFLFEDAPLSAHLKRAHFRALAEIAPISIKWNPILTEVTF
ERDVYVCGGGSFQKALTOYIADLQKVVYQVQESVGAIVGAIVICNETFOLEEMS
ANYVIEPKDOIEALAYEEMKOTQRFSGSESVLV"
complement(4036. .5070)
/gene="yoaD"
/complement(4036. .5070)
/gene="yoaD"
/translation="similar to B.subtilis D-3-phosphoglycerate
dehydrogenase (419 aa) encoded by serA gene"
/codon_start=1
/translation="yoaD"
/product="yoaD"
/protein_id="AAB84446.1"
/db_xref="PID:g2619022"
/db_xref="GI:2619022"
/translation="MKNYMKRMCSSMTVLTAPYNERGKLENLFGSVAYQSKKEG
RAYREDELQILKATNATGILTELDOYDTSFVAPLSPVGVCRGMPSNVDYAAKKE
RGJFVFTPGNMAQVAEMFIGNVLSLRHSASNMWLDKGEWMSDYLAQVYFAKKE
LTGKTVMIGFAGVORIAKLILFCDKIKYDVIQDDHPLVKAELKTVESDSDIV
SVHLPRTEELTGLIDROYFDLMKESALFVMTSVYVYVNRDDLFLVLAKEHISGAILDY
FYHPPESDVELISLPNVLATPRLAATVEVEDHVTIILKALKKRGKERTLNIQTM
```

```
gene
CDS
YNRDALKTGG"
5394. .7427
/gene="yoaE"
5394. .7427
/gene="yoaE"
/translation="similar to B.subtilis yoaE protein (667 aa) and
E.coli formate dehydrogenase (715 aa)"
/codon_start=1
/translation="putative formate dehydrogenase"
/product="putative formate dehydrogenase"
/protein_id="AAB84420.1"
/db_xref="PID:g2618996"
/db_xref="GI:2618996"
/translation="MRSFATQOQGIKFSVCSLDCPPCGGLIHKKDKGIYKVOGDDPH
PTAGNICKNRPMWTERIYDEKRLTTPDKRTGKGAIFPISMKELIDITSRMQL
IDEGAESLIPFSGNMGKLTAEGRDREFYMGSSOLERTICSRAGSGRYVTGTI
SAGIDPEETHYTHLFIWGINAVSTNHOITLQAKRKAKIIVIDVHKNOTGLAD
WFTPIKGTDSALALGIMHILEKENLHDESEVYVGEELREHVKQYDPEKVSITF
GVSTEDILRLAKYGETSPFIRINGPOHNDGMIVRTIACLPAITGQMLTGGGA
IKNSGILEYNTNALQRPDLKGTFRSFNMOLGVLETDPIRSLFIYGYNPVAV
APANKYRQGLREDLEFVYHDLFETEAADIVPATSAFENTPEYTSYHMYIOL
QOPVIERYGEKSTNTEFRLAEMGFTOELDSDPEVITROLDHPDNHIAEIIDY
SLRKHSMAKRRKRPLEPGLTPPSKTEIYSKRMQODFPALPTYPTLVTDNHEPDM
TVGPNHNFINSTFSNNEKHITLKEPTKLEIFNTRDKERHGVGAVPRLVNSGECCL
TAAVGQVLGVVVSQGLWADEQKRTACKRANRSLVRRGRHLFFRCPTD"
complement(9379. .11027)
/gene="yoaH"
/complement(9379. .9406)
/gene="yoaH"
/complement(9423. .11027)
/gene="yoaH"
/translation="similar to B.subtilis methyl-accepting chemotaxis
protein McpA (661 aa)"
/codon_start=1
/translation="yoaH"
/product="yoaH"
/protein_id="AAB84447.1"
/db_xref="PID:g2619023"
/db_xref="GI:2619023"
/translation="MISSYKKRKEELKDKMEFOKEIKRHOYEITGLSNDERGFLTRD
KEYDGMKGRADPVLSIDRVNLDLDEKQVSDIEKISFTYRRLNOQVYVAYKSN
PKAERTHGEERTIRKEGAPVAVNKLSDRDEVDLDEIDGKMSGLITIVTG
ISVILGIVLSIMLKSIMVPLRSINQLEIAEGADLTKVYVKKRDEFGOLAQSFN
SFTHSLTOIVKQILSSSEQVAASEBLSAASEKSTSEHISRAMOMADSNYKQSM
TEGASASITELDISSVASNTINDLSSMRDKAEISGVNKLMDQKFIDKQYD
SANGNLOTIVASAEIISDLSLITISEQTNILALAAIEAARAGRGGFAYAAEYV
RKADRTNSANHIQSVATIQNESIETVNNIKVQENNSGVLVSEQETGNNELN
LVQVTSQIDEVAAATQQLSGVEVLIQHTVHTLAAGTKETSANTEAVANSQQLHSM
GELSYAESLSQAEELQVYNFKY"
11755. .12910
/gene="yoai"
/citation(3)
11755. .12882
/gene="yoai"
/translation="similar to H.influenzae HI0035 protein (551 aa)"
/codon_start=1
/translation="yoai"
/product="transcription antiterminator"
/protein_id="AAB84421.1"
/db_xref="PID:g2618997"
/db_xref="GI:2618997"
/translation="MNTVMMSFASSELKDKENCFPEHILDWYEQAAKHDSLSTHTF
```

Query Match 11.0%: Score 17; DB 1; Length 87500;  
Best Local Similarity 100.0%: Pred. No. 16;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 43 CTGAATCAGAAATATC 59  
|||||  
Db 70050 CTGAATCAGAAATATC 70034





```

/EC_number="2.3.2.2"
/function="glutathione metabolism"
/note="alternate gene name: pac"
/codon_start=1
/transl_table=11
/product="gamma-glutamyltranspeptidase"
/protein_id="CAB13734.1"
/db_xref="PID:e1185314"
/db_xref="PID:g2634235"
/db_xref="GI:2634235"
/db_xref="SMISS-PROT:P54422"
/translation="MKRTWYCTALSLVLLVAGSVPEHAAKPKPSYDEKQVDY
KDGAVATDPRASEGADYKKGNSAIDAIAIOPALNTPEPMSSGIGGSPFAYVYG
KTQVTTIDSKERAPAGTIPDFEIDENKGLPESRYKGLAAGVGLKGLERLADK
WGRSMQOLITPISIKLAKGFPIDSVLAELASIDYOEKLSRTAAKDVLPNEPELKD
TLIQDLAKTFLKLSKGTDAFYKGFATLSIDYVDFGSGMTERKLENIYITDEPI
WGDYQJOIATPPSSGIFLQMLKILDFHNLSDYDRSMERYOLLAETMHSYAD
RASYAGDEFPVPLKGLHPDYIKERQOLINDQVKKPKAGDPKMYQESAMYSKY
EOPKDYBGQTHFTVADRWGNVSYTTTIGQLGTGIMVDPYCYILLNELTDALP
GGANVOPNKRPLSSMTPTILFKDKRPVLTVGSPGATISSVLQTLVHLEYMEIK
AAVEPRITYTNSMSRYREDGVPDYLSKLNGMCHKGTSPVDIGNVOSISIDHENG
FKGYADSSRNCAAGINLKRK"
/complement(5535..6484)
/gene="yofA"
/complement(5535..5559)
/gene="yofA"
/complement(5627..6484)
/gene="yofA"
/function="unknown"
/note="similar to transcriptional regulator (LysR family)"
/codon_start=1
/transl_table=11
/protein_id="CAB13735.1"
/db_xref="PID:e1185315"
/db_xref="PID:g2634236"
/db_xref="GI:2634236"
/db_xref="SPIREMBL:O34685"
/translation="MESGDLKIFQAVAREGSIITRAQMLNVQSNVTAARVHNEEDLN
IRLPHRTNRGKMLTAAGENLILQYADVLSLDQAEKSTRMSRQKGPLRTGSLTMV
THLPEHASFLRPEVDLSYNTADTHLILQOVLDHRYDGAFFVGPVEHAARQLHVS
HDEVLILSSREGTAEDMLQCPMLFEAGCSHRDKRLLEAGIHNOKIIEFGTLEAI
IKGVASMGNTALPKSAVDSEHRTNWMIHLPSPYDLEIVFYRRDFITSAFQIF
LDEINEMKR"
/6612..7601
/gene="yofA"
/6612..7601
/gene="yofA"
/function="unknown"
/note="similar to alcohol dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB13736.1"
/db_xref="PID:e1185316"
/db_xref="PID:g2634237"
/db_xref="GI:2634237"
/db_xref="SPIREMBL:O35017"
/translation="MKAVIHNGKAGLLGLSVQDVPSTRPGYGEYKVKLAGLNHRDL
FLMKSEODPHMILGSDAGIIEIEIGVKNVYQVEVYIFPLNMDLLENVPVPE

```

```

Query Match 11.0%; Score 17; DB 1; Length 207730;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 43 CTGAATCAGAAATATC 59
|||||
Db 89141 CTGAATCAGAAATATC 89125

```

```

RESULT 14
AE001272/c 60232 bp DNA circular BCT 11-SEP-1998
DEFINITION Lactococcus lactis DPC3147 plasmid pMRC01, complete plasmid
sequence..

```

```

ACCESSION AE001272
NID 93582195
VERSION AE001272.1 GI:3582195
KEYWORDS
SOURCE
ORGANISM
Lactococcus lactis.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Lactococcus.
REFERENCE
1 (bases 1 to 60232)
Dougherty, B.A., Hill, C., Weidman, J.F., Richardson, D.R., Venter, J.C.
and Ross, R.P.
Sequence and analysis of the 60 kb conjugative,
bacteriocin-producing plasmid pMRC01 from Lactococcus lactis
DPC3147
Mol. Microbiol. 29 (4), 1029-1038 (1998)
99000510
JOURNAL
MEDLINE
2 (bases 1 to 60232)
Dougherty, B.A., Hill, C., Weidman, J.F., Richardson, D.R., Venter, J.C.
and Ross, R.P.
Direct Submission
Submitted (11-SEP-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
FEATURES
location/Qualifiers
1..60232
/organism="Lactococcus lactis"
/strain="DPC3147"
/db_xref="taxon:1358"
/plasmid="pMRC01"
134..1057
/gene="ORF00001"
134..1057
/gene="ORF00001"
/note="similar to GB:L28754 PID:239026 percent identity:
71.43; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="156770 transposase"
/protein_id="AAC55992.1"
/db_xref="PID:g3582196"
/db_xref="GI:3582196"
/translation="MIESYLOHNRKPEIARNGRAIQIYVNVNFKOGKALDYWH
OYENKTKGCRKIOLPAHEVDYIKKVTIGTPDYIIGKREPVSGCRRITLYRLSK
GIKDITLPRKGRKRNKRGKRGKQYQSGSHDRDNPDPNPFSEGHLEGTIVYIH
HKSAVITLVERKSVIITIKPNRKALDITLQWFSRPKPFKSTFDCGKFSN
WKAISNQHDIDIFYADGTPSORPLNENSGILRRGKLPKSMDFREVNQTFISSVNO
RNHIPKRSILNRYRPIEFILSYVOEAFYSNLI"
/complement(1099..1365)
/gene="ORF00003"
/complement(1099..1365)
/gene="ORF00003"
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=11
/product="L. lactis predicted coding region ORF00003"
/protein_id="AAC56027.1"
/db_xref="PID:g3582231"
/db_xref="GI:3582231"
/translation="MSNOLENLISKKEIOLKIERENILKSKYLESTKKRKERTRK
LIOKGLALDKYPTENLSIDETDFLKITSNYIKENKPKYKKN"
1143..1337
/gene="ORF00002"
1143..1337
/gene="ORF00002"
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=11
/product="L. lactis predicted coding region ORF00002"
/protein_id="AAC56026.1"
/db_xref="PID:g3582230"
/db_xref="GI:3582230"
/translation="MKRFLKNLLFROLKDPQYONITLVVHPVLIFFELFFFLYOG

```

```
gene
CDS
    complement(1378..1662)
    /gene="ORF00004"
    complement(1378..1662)
    /gene="ORF00004"
    /note="hypothetical protein; identified by Genemark; putative"
    /codon_start=1
    /transl_table=11
    /product="L. lactis predicted coding region ORF00004"
    /protein_id="AAC56028.1"
    /db_xref="PID:g358232"
    /db_xref="GI:3582232"
    /translation="MTLSLEKQOEKLNLEKIKNEKKIDNKGHEIITKISNDYSKITDRIKELTEFSEFLKNGSDTERKSNDSVDENINSVHEKTENNF"
    /gene="ORF00005"
    1931..3973
    /gene="ORF00005"
    1931..3973
    /note="similar to GP:1245474 percent identity: 28.49; identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="nicking enzyme (trac)"
    /protein_id="AAC55893.1"
    /db_xref="PID:g3582197"
    /db_xref="GI:3582197"
    /translation="NAIFHMNFSNISAGGRSAVASASYSGEKLTYSEMENTTYEYNSVPESEFLLPENAEPMARDKOLMNEVAVDRKVSRAKEFNVALPIELSDDECKELTEEVOKIFGVADVAIHHDHDEPHAVHMLNRPFNADGSGOKAKVEYIDE NNGKTYAGHARSRIYLVDMDKGVKVEKRAVDHNSVPOENIDIRSEKILE AGCINDIATQHVGTGNRDERAEFKLVLENKQKALENDEKINNELKVAQKAEY SENENKVALELSKELEATFIDLEHLEKKNMFLNMKNSVLKQIVGKDVSEELKVSRO EESLDNANRLIDKVVDRSIKAMPYVDSNAFSPKPKOLIRESEKNGYETGAELDR LEMIRADSLNROIVALTRPTSLTFLMSOEKHYIDINNVLAYOGYENKEMSHGK ILRNESESEKELDIORSIKOINSIEVRKIVEQOVSVALGITPDESLKLSIVE KERTYNAVYVNPETIKKIQOSELENVYKREPILFSTKEHNOGSIAGIRREDINK HTRIVKPEGTAKKLFGEAKADKNIDPELLKENTNNDKOSTKNEYTRKNMNETOVS AVSDTSPAEYMRKLFSGVLISVLNNENOHKLEBEVDMERKHEHRTGSLR"
    4058..4402
    /gene="ORF00006"
    4058..4402
    /gene="ORF00006"
    /note="hypothetical protein; identified by Genemark; putative"
    /codon_start=1
    /transl_table=11
    /product="L. lactis predicted coding region ORF00006"
    /protein_id="AAC56029.1"
    /db_xref="PID:g3582233"
    /db_xref="GI:3582233"
    /translation="MHLKSEWHSLSKASEIILNGTSYVSLMRHEKMDPAGVNAS GSKSLIDGIEIMIKNTRKEGVLVNKRAPBDKYLILPLDKALNNYTFPKNLSIL LIYSVTLKVR"
    4403..5008
    /gene="ORF00007"
    4403..5008
    /gene="ORF00007"
    /note="hypothetical protein; identified by Genemark; putative"
    /codon_start=1
    /transl_table=11
    /product="L. lactis predicted coding region ORF00007"
    /protein_id="AAC56030.1"
    /db_xref="PID:g3582234"
    /db_xref="GI:3582234"
    /translation="MKSEKMGISGTLIHDERKGIITDKNEKSDSLDYSEKLTDKD PIKEVREKIKDSIKRDLKINPLHLKAMPDKYDSNKSKEINSDKPTLDYKOIKS DISFGESEFLGELGYFELDNVARSYSNQLIETKELGIDDEAKYFLGTSORGEF KRAISELPKRSIAEELORFSEKKEKOVOTOSIELTKDDE"
    5021..5365
    /gene="ORF00008"
    5021..5365
```

```
gene
CDS
    /gene="ORF00008"
    /note="similar to GB:L11998 PID:310609 PID:405561 percent identity: 35.58; identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="trsb protein (trab)"
    /protein_id="AAC55994.1"
    /db_xref="PID:g3582198"
    /db_xref="GI:3582198"
    /translation="MNKLGKGVAKIATSTAGYLMWMTAQTAPADGTATOGKLT TANTIKGILTGIVLVGVVALFIIRKLPDADDERKSEVRSKSVRAVAGVALAAI VMLPWSYSLT"
    5386..5733
    /gene="ORF00009"
    5386..5733
    /gene="ORF00009"
    /note="similar to GB:L11998 PID:310610 PID:405562 percent identity: 28.21; identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="trsc protein (trac)"
    /protein_id="AAC55995.1"
    /db_xref="PID:g3582199"
    /db_xref="GI:3582199"
    /translation="MKKGEFFPENVDKVGIMWDYTLKDFGILTLGCLVFIYL PYPPLFVGLKIIIVIAWNTVMALITIKPISSRNKIKVRDLRLRKYSHQKLYYL KQKGIKNEPEK"
    5717..6370
    /gene="ORF00010"
    5717..6370
    /gene="ORF00010"
    /note="similar to GB:L11998 PID:310611 PID:405563 percent identity: 33.80; identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="trsd protein (trac)"
    /protein_id="AAC55996.1"
    /db_xref="PID:g3582200"
    /db_xref="GI:3582200"
    /translation="MNLKRSKSKASALDMETIPPKINGEKETIDMSLIVGRDYOV IKTGVLGILEVSGINLIDLNDNEQDVFNDFGATLMSLGGVDDTQFIETPIPVN MTAYINGLKRKYLDLKNHPQEQRINILASLYLDHFTTVQNSKNMTTQHLIVKVI KDRSVESDLAVANNIDERKINQVRDLLENALDFDLTAIILPQOEVLVKNLNFKG"
    6382..8400
    /gene="ORF00011"
    6382..8400
    /gene="ORF00011"
    /note="similar to GB:L11998 PID:310612 PID:405564 percent identity: 11.0%; Score 17; DB 2; Length 60232; Best Local Similarity 100.0%; Pred. No. 16; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match
DB 7142 GATCTTGTATTATCC 7126
QY 73 GATCTTGTATTATCC 89
|||||
RESULT 15
AF037430 1059 bp DNA BCT 21-APR-1998
LOCUS AF037430/c
DEFINITION Bacillus subtilis membrane bound delta 5 acyl 1lipid desaturase
ACCESSION AF037430
NID 93064242
VERSION AF037430.1 GI:3064242
KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
```

REFERENCE 1 (bases 1 to 1059)  
 AUTHORS Aguilar,P.S., Cronan,J.E. Jr. and de Mendoza,D.  
 TITLE A *Bacillus subtilis* gene induced by cold shock encodes a membrane phospholipid desaturase  
 JOURNAL J. Bacteriol. 180 (8), 2194-2200 (1998)  
 MEDLINE 98215189

REFERENCE 2 (bases 1 to 1059)  
 AUTHORS Aguilar,P., Cronan,J.E. Jr. and de Mendoza,D.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-DEC-1997) Microbiology, Prombioe-Conicet-Fac. Cs. Bqcas. y Farm.- U.N.R., Suipacha 531, Rosario, Santa Fe 2000, Argentina

FEATURES  
 source Location/Qualifiers  
 1..1059  
 /organism="Bacillus subtilis"  
 /db\_xref="taxon:1423"  
 1..1059  
 /codon\_start=1  
 /transl\_table=1  
 /product="membrane bound delta 5 acyl lipid desaturase"  
 /protein\_id="AAC38355.1"  
 /db\_xref="PIR:G3064243"  
 /db\_xref="GI:3064243"  
 /translation="MTERTIAHKOKQLKQYAAFAQPTKNSLIQLNTFFPFGIMF  
 LAIISLDVSTLLTALTYIAGFLRITIFHDCCHOSFEKOKRINHILGFLTGVLTL  
 FPYLQWQSHSHIHATSSNLDKRGTDIMLTVEYKASRRIRLRYRNPIMFI  
 LGSITVFLITNRFNKKGARRERVTYLTNLAIVALAACCLIFGWSFLVQPIFL  
 ISSIGVWLFYVQHTFEDSYFEADENMSYVOAAYEGSSFYKLPKLOLGTNIGYHHV  
 HILSPKVPNTKLEVAHHEHPLKNVPTITLTKTSLSQSLAFRLMDEDNKQFVSFRAIKHI  
 PVSILPDSPEKOKLRKNA"

BASE COUNT 310 a 243 c 200 g 305 t 1 others  
 ORIGIN

Query Match 11.0%; Score 17; DB 2; Length 1059;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 CTGAATCGAATATATC 59  
 ||||||||||||||||  
 DB 661 CTGAATCGAATATATC 645

Search completed: August 6, 1999, 11:42:36  
 Job time: 6642 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:21 ; Search time 3075.15 Seconds  
(without alignments)  
159.078 Million cell updates/sec

Title: US-09-049-696-14

Perfect score: 248  
Sequence: 1 ACCTGAGCGCGAATTCAC.....TTGAAATGCGACAGATCTT 248

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	248	100.0	734	49	A1660957	A1660957 wf20d08.x
2	237	95.6	310	31	AA297150	AA297150 EST112734
C 3	214.4	86.5	501	51	AT721275	AT721275 as82h08.x
C 4	213.4	86.0	716	49	A1660234	A1660234 we68g02.x
5	149.4	60.2	629	34	AA518112	AA518112 v123c06.x
6	149.4	60.2	554	36	AA623196	AA623196 v114d07.x
7	149.4	60.2	507	36	AA623217	AA623217 v114d07.x
8	149.4	60.2	646	37	AA691335	AA691335 vs14h06.x
9	143	57.7	271	30	AA245498	AA245498 my51e11.x
10	142	57.3	552	37	AA688953	AA688953 vs04h07.x
11	125.8	50.7	370	37	AA711143	AA711143 vt36d09.x
12	123.4	49.8	301	37	AA689856	AA689856 vt62c11.x
13	119.4	48.1	441	26	W41083	W41083 mc41c10..r1
14	106.6	43.0	297	40	AA929159	AA929159 vq58g07.x
15	100.4	40.5	378	31	AA296955	AA296955 EST112726
16	97.4	39.3	482	37	AA711228	AA711228 vt70c11.x
17	94.4	38.1	452	39	AA822200	AA822200 vp36d05.x
18	88.6	35.8	243	26	W20635	W20635 mc16f08..r1
19	88.6	35.7	488	37	AA733415	AA733415 vt75b06.x
20	78.8	31.8	449	38	AA759816	AA759816 vt81b01.x
21	74.4	30.0	524	39	AA871630	AA871630 vq39a03.x
C 22	72.6	29.3	205	50	A1675991	A1675991 wd08g07.x
23	71.8	29.0	455	29	AA174903	AA174903 ms78e05.x
24	68.6	27.7	468	49	A1640041	A1640041 ms67c10.y
25	68.2	27.5	207	39	AA869318	AA869318 vq50e03.x
26	68.2	27.5	297	49	A1641931	A1641931 vq50e03.y
C 27	67.4	27.2	335	51	AT721121	AT721121 as73d08.x
28	60.8	24.5	463	37	AA690350	AA690350 vt31a09.x
C 29	60	24.2	533	33	AA429919	AA429919 zw66f05.s
30	53.8	21.7	466	29	AA170458	AA170458 ms67c10.r
31	53.6	21.6	450	29	AA182290	AA182290 mt76a07.x
C 32	46.8	18.9	175	47	A1506262	A1506262 vq50e03.x
33	41.6	16.8	289	29	AA178274	AA178274 mt11g07.x
C 34	41	16.5	488	33	AA430055	AA430055 zw65h07.s
35	39.2	15.8	431	30	AA238284	AA238284 my34g09.x
36	38.8	15.6	345	31	AA304469	AA304469 EST117232
37	38.4	15.5	469	36	AA611564	AA611564 vq85d05.x
38	37.6	15.2	401	29	AA183456	AA183456 mt26f03.x
39	37.6	15.2	418	35	AA596289	AA596289 vq26a08.x
C 40	37	14.9	387	38	AA752923	AA752923 97AS0660
41	33.6	13.5	487	26	W24316	W24316 zbd1a11..r1
42	33.2	13.4	463	39	AA839323	AA839323 vq40f07.x
C 43	32.8	13.2	744	29	AA177829	AA177829 mt28g09.x
C 44	32.6	13.1	374	31	AA278521	AA278521 zsf76b06.x
45	32.2	13.0	299	35	AA566843	AA566843 zf-A329.z

#### ALIGNMENTS

RESULT 1  
LOCUS A1660957 734 bp mRNA EST 10-MAY-1999  
DEFINITION wf20d08.x1 Soares\_Dieckgraefe\_colon\_NHUC Homo sapiens cDNA clone  
IMAGE:2351151 3' similar to TR:088826 088826 GOB-5 PROTEIN. ;, mRNA  
sequence.  
ACCESSION A1660957  
NID g4764540

VERSION A160957.1 GI:4764540  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 734)  
 AUTHORS NC1-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Mar 20, 1998 this sequence version replaced gi:2980033.

CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([infoimage.lnl.gov](http://infoimage.lnl.gov)) for further information.  
 Seq primer: -400p from Gibco  
 High quality sequence stop: 477.  
 Location/Qualifiers  
 1. 734  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="x"  
 /clone="IMAGE:2351151"  
 /clone\_lib="Soares\_Dieckgraebe\_colon\_NHUC"  
 /tissue\_type="colonic mucosa from 5 ulcerative colitis  
 patients"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - 01190(dT) primer [5'  
 TGTATCCATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3']  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization. Tissue samples  
 provided by Dr. Brian Dieckgraebe (Washington University,  
 dieckgrb.wustl.edu); colonic mucosa represents a range of  
 disease involvement from mild cryptitis to severe  
 ulceration, fibrosis, and degeneration. Library  
 constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 200 a 143 c 159 g 228 t 4 others  
 ORIGIN

Query Match 100.0%; Score 248; DB 49; Length 734;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-58;  
 Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTGAAGCGGAATTCACGGGGGAGTCTCATTAATCTGACCTGGACGCTCCTGGGG 60  
 DB 520 ACCTGAAGCGGAATTCACGGGGGAGTCTCATTAATCTGACCTGGACGCTCCTGGGG 461  
 OY 61 ATGATTAGACATGAGACAGCTCAAGTATATCATTCGATTAAGTCAATATCTTG 120  
 DB 460 ATGATTAGACATGAGACAGCTCAAGTATATCATTCGATTAAGTCAATATCTTG 401  
 OY 121 ATCTGAGAGACAGTTCATGATCTCTTCAGTGAATTAATCTGCTCATCCCAAGG 180  
 DB 400 ATCTGAGAGACAGTTCATGATCTCTTCAGTGAATTAATCTGCTCATCCCAAGG 341  
 OY 181 AAGCACTCTGAGGAGCTTTTGTGTTAAACGAGAAAACATTACTTTGAAAATGGCA 240  
 DB 340 AAGCACTCTGAGGAGCTTTTGTGTTAAACGAGAAAACATTACTTTGAAAATGGCA 281  
 OY 241 CAGATCTT 248  
 DB 280 CAGATCTT 273

RESULT 2

AA297150  
 LOCUS 310 bp mRNA EST 18-APR-1997  
 DEFINITION Colon 1 Homo sapiens CDNA 5' end, mRNA sequence.  
 ACCESSION AA297150  
 NID g1949524  
 VERSION AA297150.1 GI:1949524  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 310)  
 AUTHORS Adams,M.D., Kerlavage,A.R., Flieschmann,R.D., Fuldner,R.A.,  
 But,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D.,  
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Mai,C.,  
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
 Glodex,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,  
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudex,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Uteback,T.R., Weidman,J.F., Li,Y.,  
 Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
 Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
 He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
 Korak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,  
 Raymond,D., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
 Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
 Fraser,C.M. and Venter,J.C.  
 Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl), 3-174 (1995)  
 9602580  
 On Sep 12, 1996 this sequence version replaced gi:1288566.  
 COMMENT OTHER ESTs: THC167738  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: [arkerlav@tigr.org](mailto:arkerlav@tigr.org)  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.cigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1. 310  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (lnhost):127942"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Colon 1"  
 /dev\_stage="adult"  
 /note="Organ: colon; Vector: pbluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI"

BASE COUNT 97 a 61 c 58 g 91 t 3 others  
 ORIGIN

Query Match 95.6%; Score 237; DB 31; Length 310;  
 Best Local Similarity 99.2%; Pred. No. 3.8e-55;  
 Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 10 CGGAATTCACGGGGGAGTCTCATTAATCTGACCTGGACAGCTCTGGGATATTATG 69  
 DB 1 CGGAATTCACGGGGGAGTCTCATTAATCTGACCTGGACAGCTCTGGGATATTATN 60  
 OY 70 ACCATGGAACGCTCACAGATATATCATTCGATTAAGTCAAGATATCTGATCAGAG 129  
 DB 61 ACCATGGAACGCTCACAGATATATCATTCGATTAAGTCAAGATATCTGATCAGAG 120  
 OY 130 ACAAGTCAATGATCTCTCAAGTGAATTAATCTGCTCATCCCAAGAGCACT 189

Db 121 ACAAGTCATGATATNTTCAAGTACATCTGCTCTCATCCCAAGAGCACT 180  
OY 190 CTGAGGAGCTCTTTTGTGTTAAACGAGAAACATTTGTAATAATGGCAGATCTT 248  
|||||  
Db 181 CTGAGGAGCTCTTTTGTGTTAAACGAGAAACATTTGTAATAATGGCAGATCTT 239  
|||||  
RESULT 3  
A1721275 501 bp mRNA EST 10-JUN-1999  
A1721275/c as82h08.x1 Bartshead colon HPLRB7 Homo sapiens cDNA clone  
LOCUS IMAGE:2335263 3' similar to SW:ECIC\_BOVIN P54281 EPITHELIAL  
DEFINITION CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.  
ACCESSION A1721275  
NID 95038531  
VERSION A1721275.1 GI:5038531  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
REFERENCE Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 501)  
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Getzel,G., Jost,S.,  
Kritman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marr,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
TITLE WashU-NCI human EST Project  
JOURNAL On Jun 22, 1998 this sequence version replaced gi:3246918.  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from Gibco  
High quality sequence stop: 394.  
FEATURES  
Location/Qualifiers  
1..501  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="Xp11.2"  
/clone="IMAGE:2335263"  
/clone\_1ib="Bartshead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: pTR73D-Pac (Pharmacia) with a  
modified polylinker; Site:1: EcoRI; Site:2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dt) primer [5'  
TGTTACGATCTGAGTGGAGCGGCGCCCTTTTGTGTTTGTGTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[5', AATCACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested  
with Not I and cloned into the Not I and Eco RI sites of  
the modified pTR73 vector. Library constructed by Bob  
Bartshead."  
BASE COUNT 158 a 83 c 106 g 154 t  
ORIGIN  
Query Match 86.5%; Score 214.4; DB 51; Length 501;  
Best Local Similarity 99.5%; Pred. NO. 5.5e-49;  
Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 33 ATTAATCTGACTTGACAGCTCCTGGGATGATATGACCATGGAACAGCTCAAGATAT 92  
|||||  
Db 501 ATTAATCTGACTTGACAGCTCCTGGGATGATATGACCATGGAACAGCTCAAGATAT 442  
|||||  
OY 93 ATCAATGGAATGAAGTACAGTATTTGATCTCAGAGCAAGTTCAATGATCTCTTCAA 152  
|||||

Db 441 ATCAATGGAATGAAGTACAGTATTTGATCTCAGAGCAAGTTCAATGATCTCTCAA 382  
OY 153 GTGATATGACTGCTGCTCATCCCAAGAGCAACTGTGAGGAGCTCTTTGTTAA 212  
|||||  
Db 381 GTGATATGACTGCTGCTCATCCCAAGAGCAACTGTGAGGAGCTCTTTGTTAA 322  
|||||  
OY 213 CCAGAAACATTTACTTTGAAATGGCAGATCTT 248  
|||||  
Db 321 CCAGAAACATTTACTTTGAAATGGCAGATCTT 286  
|||||  
RESULT 4  
A1660234 716 bp mRNA EST 10-MAY-1999  
A1660234/c we68g02.x1 Soares.Dieckgraefe.colon\_NHCD Homo sapiens cDNA clone  
LOCUS IMAGE:2346290 3' similar to TR:088826 088826 GOB-5 PROTEIN. ;, mRNA  
DEFINITION sequence.  
ACCESSION A1660234  
NID 94763804  
VERSION A1660234.1 GI:4763804  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
REFERENCE Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 716)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949219.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 387.  
FEATURES  
Location/Qualifiers  
1..716  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2346290"  
/clone\_1ib="Soares.Dieckgraefe.colon\_NHCD"  
/tissue\_type="colonic mucosa from 3 patients with Crohn's  
disease."  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pTR73D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dt) primer [5'  
TGTTACGATCTGAGTGGAGCGGCGCCCTTTTGTGTTTGTGTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pTR73 vector. Library  
provided through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieck@im.wustl.edu); colonic mucosa represents a range of  
disease involvement from moderate to severe Crohn's  
disease; samples include both perforating (fistulas) and  
non-perforating samples. Library constructed by Bento  
Soares and M. Fatima Bonaudo.  
BASE COUNT 220 a 123 c 140 g 233 t  
ORIGIN  
Query Match 86.0%; Score 213.4; DB 49; Length 716;  
Best Local Similarity 93.3%; Pred. NO. 1e-48;  
Matches 223; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
OY 10 CGGAATTCACGGGGGAGCTTCATTAATCTGACTTGACAGCTCCTGGGATGATATG 69  
|||||

**FEATURES**

**SOURCE**

Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 351.

Location/Qualifiers  
1. .629

```
/organism="Mus musculus"
/strain="FVB/N"

```

```
/ab_xref="taxon:10090"  
/clone="IMAGE:904618"  
/clone_lib="Barstead mouse irradiated colon MPI.RB7"
```

```
/dev_stage="8 weeks"
/lab_host="DH10B"
```

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained from 8 newborn mice. Cells were harvested 70 hours after

irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - oligo(dT) primer

[5',TGTTACGATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (AATTCGATCCTTG), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector library constructed by Bob Barstead.

BASE COUNT  
ORIGIN

	RESULT	6			
AA623196	LOCUS				
DEFINITION		AA623196	554 bp	mRNA	EST
		V114B07.1	Bairstead mouse proximal colon MPELB6	Mus musculus	CDNA
		clone IMAGE:903733	5' similar to F8:G1184066	G1184066	
		CALCIUM-ACTIVATED CHLORIDE CHANNEL.	/; mRNA sequence.		
ACCESSION		AA623196			
NID		Q2527072			
VERSION		AA623196.1	GI:2527072		
KEYWORDS		EST.			
SOURCE		house mouse.			
ORGANISM		Mus musculus			

**SOURCE ORGANISM**

## REFERENCE

## AUTHORS

TITLE  
JOURNAL

COMMENT

## FEATURES

**SOURCE**





Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:619507  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 491.

## FEATURES

## SOURCE

1. 646  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /map="17q21. 2"  
 /clone="IMAGE:1138235"  
 /clone\_lib="Barstead mouse irradiated colon MPLRB7"  
 /dev\_stage="8 weeks"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained  
 from 8 week old mouse. Colon was harvested 72 hours after  
 irradiation with 1400 Gys. 1st strand cDNA was primed  
 with a Not I - oligo(dT) primer  
 [5'TGTACGATCTGAGTGGAGGCGGCCCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors [AATTCGATCCTTG], digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT73  
 vector. Library constructed by Bob Barstead."  
 BASE COUNT 171 a 176 c 153 g 146 t  
 .ORIGIN

Query Match 60.2%; Score 149.4; DB 37; Length 646;  
 Best Local Similarity 75.3%; Pred. No. 2.8e-31;  
 Matches 186; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Oy 1 ACCTGAAGCGGGAATTCACGGGGGACGTCATTAATCTGACCTGGAGAGCTCCTGGGG 60  
 |||||  
 Db 201 ACCTGAAGCGGGAATTCACGGGGGACGTCATTAATCTGACCTGGAGAGCTCCTGGGG 260  
 Oy 61 ATGATTATGACCATGGAACAGCTCAAGTATATCATTCGATTAAGTACAGTATCTTG 120  
 |||||  
 Db 261 ATGATTATGACCATGGAACAGCTCAAGTATATCATTCGATTAAGTACAGTATCTTG 320  
 Oy 121 ATCTCAGAGCAAGTTCATGATCTCTTCAAGTAACTACTGCTCATCCCAAG 180  
 |||||  
 Db 321 ATCTCAGAGCAAGTTCATGATCTCTTCAAGTAACTACTGCTCATCCCAAG 380  
 Oy 181 AAGCAACTCTGAGAGCTTTTGTGTTAAACCAAGAAACATTTGAAATGGCA 240  
 |||||  
 Db 381 AAGCAACTCTGAGAGCTTTTGTGTTAAACCAAGAAACATTTGAAATGGCA 440  
 Oy 241 CAGATCT 247  
 |||||  
 Db 441 CAGATAT 447

RESULT 9  
 LOCUS AA245498 271 bp mRNA EST 10-MAR-1997  
 DEFINITION myt1ell.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA  
 CLONE IMAGE:699404 5' similar to TR:61184066 G1184066  
 CALCIUM-ACTIVATED CHLORIDE CHANNEL.; mRNA sequence.  
 ACCESSION AA245498  
 NID 91876284  
 VERSION AA245498.1 GI:1876284  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 271)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisels, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Apr 14, 1993 this sequence version replaced g1:693201.

## COMMENT

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:432964

Putative full length read

vector to vector length is 658

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 134.

## FEATURES

## SOURCE

1. 271  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /map="14"  
 /clone="IMAGE:699404"  
 /clone\_lib="Barstead mouse pooled organs MPLRB4"  
 /sex="mixed"  
 /tissue\_type="pooled organs"  
 /dev\_stage="7 day"  
 /lab\_host="DH10B"  
 /note="Organ: pooled. Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTACGAATCTGAGTGGAGGCGGCCCTTTTCTTTTCTTTTCTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 [GTGGAATCGGATC], digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 Library constructed by Bob Barstead."  
 BASE COUNT 74 a 77 c 65 g 55 t  
 .ORIGIN

Query Match 57.7%; Score 143; DB 30; Length 271;  
 Best Local Similarity 73.7%; Pred. No. 1.6e-29;  
 Matches 182; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Oy 1 ACCTGAAGCGGGAATTCACGGGGGACGTCATTAATCTGACCTGGAGAGCTCCTGGGG 60  
 |||||  
 Db 24 ACCTGAAGCGGGAATTCACGGGGGACGTCATTAATCTGACCTGGAGAGCTCCTGGGG 83  
 Oy 61 ATGATTATGACCATGGAACAGCTCAAGTATATCATTCGATTAAGTACAGTATCTTG 120  
 |||||  
 Db 84 ATGATTATGACCATGGAACAGCTCAAGTATATCATTCGATTAAGTACAGTATCTTG 143  
 Oy 121 ATCTCAGAGCAAGTTCATGATCTCTTCAAGTAACTACTGCTCATCCCAAG 180  
 |||||  
 Db 144 ATCTCAGAGCAAGTTCATGATCTCTTCAAGTAACTACTGCTCATCCCAAG 203  
 Oy 181 AAGCAACTCTGAGAGCTTTTGTGTTAAACCAAGAAACATTTGAAATGGCA 240  
 |||||  
 Db 204 AAGCAACTCTGAGAGCTTTTGTGTTAAACCAAGAAACATTTGAAATGGCC 263  
 Oy 241 CAGATCT 247  
 |||||  
 Db 264 CCGGTTT 270

RESULT 10



from 8 week old mouse. Colon was harvested 72 hours irradiation with 1400 Gys. 1st strand cDNA was prime with a Not I - oligo(dT) primer

```

/db_xref="taxon:10090"
/map="15"
/clone="IMAGE:351090"

```

```

/comp 15
/clone="IMAGE:351090"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"

```

```

/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pT733 (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer (5'
TGTACCAATCTGTAAGTGGGACGCGCATTTTCTTTTCTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT733 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
```

Query Match	48.1%	Score	119.4	DB	26	Length	441
Best Local Similarity	73.2%	Pred	No. 4.2e-23				
Matches	153	Conservative	0	Mismatches	56	Indels	0
						Gaps	0

OY	33	ATTATCTGATTTGGACAGCTCCCTGGGGATGATATATACCATGGAAAGCTCAAGAT	92
Db	174	ATCAGTCTACATGTGAGCGGCCCCAGGAGATGATTTATGTTGGAGAGTTCAACAGTAT	233
OY	93	ATCATTGGAATTAAGTACAGATATTCTTGATCTCGAGACAAAGTTCATGATCTCTCAA	152
Db	234	ATCATAGGAACACAGCAAAATATTCATTGATCTCGAGACAAATTTATATTCCTTCGG	293
OY	153	GTGAATCTACTGCTCTTCATCCCAAGGAAGCCACTCTGAGGAGTCTTTTGTTTAA	212
Db	294	GTTATATCTACTTAAACTTACACCAAAAGAGCCCACTCAAGAAACCTTTGCCCTTAA	353
OY	213	CCAGAAACATTTACTTTTGAAGATGGAC	241
Db	354	CCAGAAATATATCTCAGAGAGAAATGCAAC	382

RESULT	14
AA929159	
LOCUS	297 bp mRNA
DEFINITION	v589ng7.1 Barsted mouse proximal colon MPlRB6 Mus musculus cDNA
IMAGE	1106556 5' similar to SM:BCLC_BOVIN P54281 EPITHELIAL
CHLORIDE CHANNEL PROTEIN ;	mRNA sequence.

ACCESSION	AA929159
NTID	93078468
VERSION	AA929159.1
KEYWORDS	GI:3078468
SOURCE	EST.
ORGANISM	house mouse.
	Mus musculus

REFERENCE  
AUTHORS  
1 (bases 1 to 297)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Gelsel, S., Kucaba, T., Lacy, M., LeM., Martin, J., Morris, M.,  
Schellendberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, B., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE  
The Washu-HMI Mouse EST Project  
JOURNAL  
Unpublished (1996)  
COMMENT  
On Jan 19, 1998 this sequence version replaced g1:2152936.

Contact: Maria M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through INL; contact the  
IMAGE Consortium (info@image.inl.gov) for further information.  
MGI:604724  
Seq primer: -20m13 rev2 ET from Amersham  
High quality sequence stop: 185.

FEATURES	SOURCE
Location/Qualifiers	l..297
/organism="Mus musculus"	
/strain="FVB/N"	
/db_xref="taxon:10090"	
/clone_image="1106556"	
/clone_id="Barstead mouse proximal colon MPLRB6"	
/dev_stage="7 day juvenile"	
/lab_host="DH10B"	
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: EcoRI. Site_2: NotI. 1st strand cDNA was primed with a Not I - oligo(dt) primer [5' TGTTACGATCTGAAGTGGAGCGCCGCTTTTTTTTTTTTTTTTTT TTTTGATTCGATCCTTGG]. double-stranded cDNA was ligated to Eco RI adaptors [AATTCGATCCTTGG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector.	
Library constructed by Bob Barstead."	
BASE COUNT	110 a
ORIGIN	62 c 50 g 75 t

[illegible]

Qy	237	GGCAC	241
Db	194	GCAAC	198

RESULT	15
AA296955	
LOCUS	378 bp mRNA
DEFINITION	EEST112726 Colon I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION	AA296955
NID	G1949515
VERSION	AA296955.1 GI:1949515

ORGANISM	TECHNIQUE
<i>Homo sapiens</i>	
Euthyrotia, Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	
1 (bases 1 to 378)	
AUTHORS	
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,	

White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Georhaegen,N.S., Godelok,A., Gneulm,C.L., Hanna,M.C., Hedblom,E., Hinkler,P.S., Jr., Kelley,J.M., Kelley,D.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palancques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,L.L., Saudak,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uteback,T.R., Weldman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrite,A., Fischer,C., Haastings,G.A., He,M.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungun,Ju,C., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,Y., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Hasseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.

Mon Aug 9 13:26:02 1999

us-09-049-696-14.rst

Page 10

JOURNAL based upon 83 million nucleotides of cDNA sequence  
MEDLINE Nature 377 (6547 suppl), 3-174 (1995)  
COMMENT 96026280  
On Sep 12, 1996 this sequence version replaced g1:1288227.  
Other\_ESTs: THC167738  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlavage@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
Location/Qualifiers  
1..378  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):127834"  
/db\_xref="taxon:9606"  
/clone\_lib="Colon I"  
/dev\_stage="adult"  
/note="Organ: Colon; Vector: pBluescript SK-; Site\_1:  
EcorI; Site\_2: XhoI"  
BASE COUNT 112 a 79 c 64 g 119 t 4 others

ORIGIN  
BASE COUNT 112 a 79 c 64 g 119 t 4 others

Query Match 40.5%; Score 100.4; DB 31; Length 378;  
Best Local Similarity 99.0%; Pred. No. 6,4e-18;  
Matches 101; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 147 CTTCAAGTGAATACACTGCTCTCATCCCAAGAGCCCACTGTGAGAGCTTTTG 206  
DB 1 CTTCAAGTGAATACACTGCTCTCATCCCAAGAGCCCACTGTGAGAGCTTTTG 60  
OY 207 TTTAAACCGAAGAAACATTTTGAATAATGGCAGATCTT 248  
DB 61 TTTAAACCGAAGAAACATTTTGAATAATGGCAGATCTT 102

Search completed: August 6, 1999, 13:33:25  
Job time: 13003 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:41:00 ; Search time 650.92 Seconds

(without alignments)  
95.323 Million cell updates/sec

Title: US-09-049-696-14

Perfect score: 248  
Sequence: 1 ACCTGAAGCGGAAATTCAC.....TTGAAATGCGACAGATCTT 248

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	248	100.0	878	1 T45884	Human colon specif
2	248	100.0	878	1 V16672	Polynucleotide seq
3	19	7.7	2100	1 V59671	Human secreted pro
4	17	6.9	420	1 O57632	Endothelin-1 (ET-1
5	17	6.9	713	1 T31013	Human endothelin-1
6	17	6.9	14507	1 T73568	Expression augment
7	17	6.9	12461	1 X32367	Human endothelin-1
8	16	6.5	5100	1 O37760	Glucosyltransferase
9	16	6.5	231	1 T19623	Human gene signatu
10	16	6.5	2139	1 T15718	Human gene signatu
11	16	6.5	2139	1 T18830	Human gene signatu
12	16	6.5	390	1 T22284	Human gene signatu
13	16	6.5	287	1 T23588	Human gene signatu
14	16	6.5	1593	1 T86242	CDNA for birch pol
15	16	6.5	16995	1 V52215	Streptococcus pneu
16	16	6.5	3175	1 V43323	Human secreted pro
17	16	6.5	3259	1 V43306	Human secreted pro
18	16	6.5	110000	1 X30248_08	Continuation (9 of
19	15	6.0	1026	1 Q03968	PENY9 encoding a r
20	15	6.0	2263	1 N90398	HIV-1 env gene seg
21	15	6.0	1059	1 N90740	DNA encoding HIV p
22	15	6.0	1791	1 N90741	DNA encoding HIV p
23	15	6.0	420	1 N91287	DNA encoding HIV-1
24	15	6.0	127	1 N91095	Epitope to HIV. He
25	15	6.0	1036	1 N92109	Synthetic polynuc
26	15	6.0	1796	1 N92110	Synthetic polynuc
27	15	6.0	2511	1 N92113	Polynucleotide enc
28	15	6.0	1920	1 N92114	Polynucleotide enc
29	15	6.0	1059	1 N92117	Polynucleotide enc
30	15	6.0	109	1 N94566	Epitope to HIV. He
31	15	6.0	2237	1 N90531	Sequence encoding
32	15	6.0	2553	1 N80949	HIV protein H76. P
33	15	6.0	1791	1 N92118	Sequence encoding
34	15	6.0	2457	1 N81451	Sequence encoding
35	15	6.0	1920	1 N81452	Sequence encoding
36	15	6.0	1803	1 N70875	HIV-III gag/env g
37	15	6.0	1803	1 N70835	Sequence encoding
38	15	6.0	1140	1 N91289	Sequence encoding
39	15	6.0	2607	1 N60077	DNA sequence of th
40	15	6.0	78	1 N60669	HIV virus env reg1
41	15	6.0	51	1 N60672	HIV virus env reg1
42	15	6.0	4020	1 N71016	Sequence of LAV/HIT
43	15	6.0	891	1 N71251	Sequence encoding

44 15 6.0 854 1 N71256  
c 45 15 6.0 3997 1 V26082

Sequence of HIV-I  
Tomato pest resist

## ALIGNMENTS

RESULT	ID	Location/Qualifiers
1	T45884	W09639419-AL.
2	T45884	12-DEC-1996.
3	T45884	06-JUN-1995; U07289.
4	T45884	06-JUN-1995; W0-U07289.
5	T45884	(HUMA-) HUMAN GENOME SCI INC.
6	T45884	Rosen CA, Yu G;
7	T45884	WPI; 97-043054/04.
8	T45884	P-PSDB; W05548.
9	T45884	Human colon specific genes and their expression products - detection of
10	T45884	of which, in non-colon tissue samples, can be used as indication of
11	T45884	colon cancer metastasis
12	T45884	Clam 1; Fig 5; 60p; English.
13	T45884	13 CDNA clones (T45880-92), most of them partial clones, correspond
14	T45884	to human colon specific genes, designated CSG1, CSG2, etc., that
15	T45884	are primarily expressed in tissues derived from the colon. CSG7
16	T45884	and CSG10 show reduced expression in colon cancer cells as compared
17	T45884	to that in normal cells; the remaining genes are overexpressed in
18	T45884	colon cancer. The partial CDNA sequences can be used to isolate
19	T45884	full-length clones and genomic clones including the complete gene.
20	T45884	CSG nucleic acids can be used to produce CSG polypeptides (see also
21	T45884	W05645-53) in transformed host cells, as probes to detect disorders
22	T45884	of the colon, partic. colon cancer and colon cancer metastasis, and
23	T45884	in gene therapy.
24	T45884	Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;
25	T45884	Query Match 100.0%; Score 248; DB 1; Length 878;
26	T45884	Best Local Similarity 100.0%; Pred. No. 1,9e-120;
27	T45884	Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
28	T45884	1 ACCTGAAGCGGAAATTCACGGGGGAGTCTCATTAATCTGACCTTGACAGCTCTCTGGGG 60
29	T45884	
30	T45884	306 ACCTGAAGCGGAAATTCACGGGGGAGTCTCATTAATCTGACCTTGACAGCTCTCTGGGG 365
31	T45884	
32	T45884	61 ATGATTATGACATGAGACAGTCAAGTATATATATGATGATGATGATGATGATGATGATG 120
33	T45884	
34	T45884	366 ATGATTATGACATGAGACAGTCAAGTATATATATGATGATGATGATGATGATGATGATG 425
35	T45884	
36	T45884	121 ATCTGAGAGACAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
37	T45884	
38	T45884	426 ATCTGAGAGACAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 485
39	T45884	
40	T45884	181 AAGCAACTCTGAGAGAGCTTTTGTGTTAAACGAGAAACATTAATCTTTGAAATGGA 240
41	T45884	
42	T45884	486 AAGCAACTCTGAGAGAGCTTTTGTGTTAAACGAGAAACATTAATCTTTGAAATGGA 545
43	T45884	
44	T45884	241 CAGATCTT 248
45	T45884	
46	T45884	546 CAGATCTT 553
47	T45884	
48	T45884	RESULT 2
49	T45884	V16672
50	T45884	Standard; CDNA; 878 BP.





22-AUG-1997: US-056872.  
PR 22-AUG-1997: US-056874.  
PR 22-AUG-1997: US-056875.  
PR 22-AUG-1997: US-056876.  
PR 22-AUG-1997: US-056877.  
PR 22-AUG-1997: US-056878.  
PR 22-AUG-1997: US-056879.  
PR 22-AUG-1997: US-056880.  
PR 22-AUG-1997: US-056881.  
PR 22-AUG-1997: US-056882.  
PR 22-AUG-1997: US-056884.  
PR 22-AUG-1997: US-056886.  
PR 22-AUG-1997: US-056887.  
PR 22-AUG-1997: US-056888.  
PR 22-AUG-1997: US-056889.  
PR 22-AUG-1997: US-056892.  
PR 22-AUG-1997: US-056893.  
PR 22-AUG-1997: US-056894.  
PR 22-AUG-1997: US-056903.  
PR 22-AUG-1997: US-056908.  
PR 22-AUG-1997: US-056909.  
PR 22-AUG-1997: US-056910.  
PR 22-AUG-1997: US-056911.  
PR 05-SEP-1997: US-057650.  
PR 05-SEP-1997: US-057659.  
PR 05-SEP-1997: US-057761.  
PR 12-SEP-1997: US-057785.  
PR (HUMA-1) HUMAN GENOME SCI INC.  
PI Bednarik DP, Brenner LA, Carter KC, Duan R, Ebner R, Endress GA,  
PI Feng P, Ferrite AM, Fischer CL, Florence KA, Greene JM, Hu JS,  
PI Kysw H, Lafleur DM, Li Y, Moore PA, NJ J, Olsen HS, Rosen CA,  
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
PI WPI: 98-506364/43.  
DR P-PSDB: W74888.  
DR  
PT New isolated human genes and the secreted polypeptide(s) they encode  
PT - useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
PT  
PS Claim 1: Page 399-400; 721pp; English.  
CC This sequence represents a nucleic acid molecule designated gene 161 from  
CC the human cDNA clone HAWBA28 (deposited as clone ATCC 97903 and ATCC  
CC 208049) which encodes a secreted human protein. The gene can be used to  
CC generate fusion proteins by linking to the gene to a human immunoglobulin  
CC Fc portion (e.g. V59502) for increasing the stability of the fused  
CC protein as compared to the human protein only.  
CC The invention relates to 186 novel genes and their fragments (nucleic  
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75025) which  
CC are useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. Also, pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the 186 polynucleotides, based on  
CC which tissues they are most highly expressed in (see V59511 for described  
CC uses).  
SQ Sequence 2100 BP; 630 A; 387 C; 406 G; 672 T;

```

Query Match          7.7%; Score 19; DB 1; length 2100;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY      60 GATGATATGACCATGGAA 78
        |||
Db       1460 GATGATATGACCATGGAA 1478

RESULT  4
ID       O57632 standard; DNA: 420 BP.
AC       O57632:
DE       17-AUG-1994 (first entry)
DE       Endothelin-1 (ET-1) promoter region.
KW       Endothelin-1; ET-1; vascular disease; cardiovascular disease;
        gene therapy; drug delivery; thrombosis; thrombolysis;

```

KW	inflammation; immunomodulation; promoter; gene expression; ss.
OS	Homo sapiens.
EH	Key
FT	cds
FT	location/Qualifiers
FT	191
FT	/*tag= a
FT	/note= "Transcriptional start site."
FT	155..158
FT	/*tag= b
FT	220..226
FT	/*tag= c
FT	103..134
FT	/*tag= d
FT	/label= Region A.
FT	/note= "Region A contains positive regulatory
FT	sequences that function in conjunction with
FT	other elements of the ET-1 promoter in a
FT	stimulatory fashion. Region A interacts
FT	with region B to induce high rates of
FT	transcription of genes under their control."
FT	122..153
FT	/*tag= e
FT	/label= Region B.
FT	/note= "Region B interacts with region A to induce
FT	high rates of transcription of genes under
FT	their control."
FT	47
FT	/*tag= f
FT	/note= "This is the 5' most nucleotide of the
FT	chloramphenicol acetyltransferase (CAT)
FT	construct 204CAT."
FT	108
FT	/*tag= g
FT	/note= "This is the 5' most nucleotide of the
FT	chloramphenicol acetyltransferase (CAT)
FT	construct 143CAT."
FT	122
FT	/*tag= h
FT	/note= "This is the 5' most nucleotide of the
FT	chloramphenicol acetyltransferase (CAT)
FT	construct 129CAT."
FT	153
FT	/*tag= i
FT	/note= "This is the 5' most nucleotide of the
FT	chloramphenicol acetyltransferase (CAT)
FT	construct 98CAT."
FT	208
FT	/*tag= j
FT	/note= "This is the 5' most nucleotide of the
FT	chloramphenicol acetyltransferase (CAT)
FT	construct 43CAT."
FT	47..165
FT	/*tag= k
FT	/note= "A 119 base pair fragment used in
FT	heterologous promoter experiments."
FT	
PN	US5288846-A.
PD	22-FEB-1994.
PF	19-OCT-1990; 598890.
PR	19-OCT-1990; US-598890.
PR	03-AUG-1992; US-924396.
PA	(GEHO) GEN HOSPITAL CORP.
PI	Lee M, Quertermos T;
PI	WPI; 94-064943/08.
PT	DNA comprising endothelin-1 promoter elements - for expression of
PT	heterologous genes in vascular endothelial cells, esp. for gene
PT	therapy
PS	Claim 1; Figure 1A; 15pp; English.
CC	The promoter element is capable of directing high-level, cell-
CC	specific expression of an operably linked heterologous gene. The
CC	promoter may be used for gene therapy and drug delivery for vascular
CC	and cardiovascular diseases by expression of proteins involved in
CC	thrombolysis, thrombolysis, inflammation, immunomodulation, etc. It
CC	may also be used in transgenic animal studies of human diseases.
CC	Sequence 420 BP: 76 A; 125 C; 109 G; 110 T;

Query Match 6.9%; Score 17; DB 1; Length 420;  
 Best Local Similarity 100.0%; Pred. No. 7.7;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 TCTGACTTGACAGCTC 54  
 ||||||||||||||||  
 DB 253 TCTGACTTGACAGCTC 237

RESULT 5  
 T31013/C  
 ID T31013 standard; DNA: 713 BP.  
 AC T31013;  
 DT 26-SEP-1996 (first entry)  
 DE Human endothelin-1 gene promoter.  
 KW Gene therapy; hypoxia related enhancer element; HREE; ischaemia;  
 KW reperfusion; promoter; endothelin-1; endothelium; ss.  
 OS Homo sapiens.  
 PN W09620276-A1.  
 PD 04-JUL-1996.  
 PE 13-NOV-1995; IB0996.  
 PR 23-DEC-1994; US-365486.  
 PA (STR1 ) SRI INT.  
 PI Bihoptric NH, Green CJ, Laderoute KR, Murphy B;  
 PI Webster KA;  
 DR WPI; 96-321849/32.  
 PT Chimeric gene conty. therapeutic gene linked to HREE - partic. for  
 PT expressing SOD etc. in hypoxic tissue to reduce tissue injury caused  
 PT by ischaemia or reperfusion  
 PS Example 1: Page 48, 118pp; English.  
 CC Expression of pGL3-1700, contg. the human endothelin-1 (ET-1)  
 CC gene promoter (T31013), in transfected human arterial endothelial  
 CC cells was increased approx. 5-fold by induction of the cells in  
 CC a hypoxic atmosphere. No hypoxia-induced increase in expression  
 CC was seen in HeLa cells, C2C12 cells or cardiac myocytes. The  
 CC promoter may be used to target hypoxia-regulated genes specifically  
 CC to cells of the vascular endothelium, since it contains elements  
 CC conferring tissue specificity as well as hypoxia response enhancer  
 CC elements able to up-regulate transcription of a (therapeutic) gene  
 CC under hypoxic conditions.  
 SQ Sequence 713 BP; 191 A; 168 C; 174 G; 180 T;

Query Match 6.9%; Score 17; DB 1; Length 713;  
 Best Local Similarity 100.0%; Pred. No. 7.7;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 TCTGACTTGACAGCTC 54  
 ||||||||||||||||  
 DB 680 TCTGACTTGACAGCTC 664

RESULT 6  
 T73568/C  
 ID T73568 standard; DNA: 14507 BP.  
 AC T73568;  
 DT 26-SEP-1997 (first entry)  
 DE Expression augmenting sequence element (EASE).  
 KW Expression augmenting sequence element; EASE; vector;  
 KW recombinant protein; gene expression; Chinese hamster ovary; CHO;  
 KW ss.  
 OS Cricetus sp.  
 FH Key Location/Qualifiers  
 FT misc\_feature  
 FT 1.14507  
 FT /tag- a  
 FT /note- "EASE (Claim 1)"  
 FT 5980.14507  
 FT /tag- b  
 FT /note- "EASE (Claim 1)"  
 FT 8671.14507  
 FT misc\_feature  
 FT /tag- c

FT /note- "EASE (Claim 1)"  
 FT 8671.10515  
 FT /tag- d  
 FT /note- "EASE (Claim 1)"  
 FT 9277.10515  
 FT /tag- e  
 FT /note- "EASE (Claim 1)"  
 FT 8672.11273  
 FT /tag- f  
 FT /note- "EASE (Claim 1)"  
 FT 10100.14923  
 FT /tag- g  
 FT /note- "EASE (Claim 1)"  
 FT 14290.14507  
 FT /tag- h  
 FT /note- "the EASE sequence is preferably ligated  
 to DNA comprising nucleotides 14290-14507  
 (Claim 3)"

PN W09725420-A1.  
 PD 17-JUL-1997.  
 PE 10-JAN-1997; U00483.  
 PR 11-JAN-1996; US-586509.  
 PA (IMMV ) IMMUNEX CORP.  
 PI Lee C, Morris AE, Thomas JN;  
 DR WPI; 97-372861/34.  
 PT Expression augmenting sequence elements - used in vectors for  
 PT expressing recombinant proteins at high levels in shorter periods of  
 PT time  
 PS Claim 1: Page 23-30; 36pp; English.  
 CC A 14507 bp DNA sequence (T73568) comprises a novel transcription  
 CC regulatory element, expression augmenting sequence element (EASE),  
 CC that facilitates high expression of recombinant proteins in  
 CC mammalian host cells. It was identified by cloning the integration  
 CC site of a unique expression cassette encoding recombinant dimeric  
 CC tumour necrosis factor receptor-immunoglobulin Fc fusion protein  
 CC from genomic DNA of a CHO 245-3 cell line expressing this protein  
 CC at a high level. Expression vectors incorporating an EASE show  
 CC a 2- to 8-fold improvement of recombinant protein expression levels  
 CC in CHO host cells. The EASE sequence also facilitates high  
 CC expression in shorter periods of time.  
 SQ Sequence 14507 BP; 4225 A; 2548 C; 3136 G; 4598 T;

Query Match 6.9%; Score 17; DB 1; Length 14507;  
 Best Local Similarity 100.0%; Pred. No. 8;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 AGAACAATTACTTTG 231  
 ||||||||||||||||  
 DB 12460 AGAACAATTACTTTG 12444

RESULT 7  
 X32367/C  
 ID X32367 standard; DNA: 12461 BP.  
 AC X32367;  
 DT 16-JUN-1999 (first entry)  
 DE Human endothelin-1 (ET-1) gene sequence.  
 KW Pulmonary hypertension; therapeutic; aerosolized; endothelin-1; ET-1;  
 KW lung; antisense; ss.  
 OS Homo sapiens.  
 PN W09911778-A1.  
 PD 11-MAR-1999.  
 PE 02-SEP-1998; G02584.  
 PR 02-SEP-1997; GB-018487.  
 PA (UYSH-) UNIV SHEFFIELD.  
 PI Higebottam T, McCormack K, Smith A;  
 DR WPI; 99-205185/17.  
 PT New composition containing an aerosolized antisense ET-1 molecule -  
 PT useful for treating pulmonary hypertension  
 PS Disclosure; Fig 1: 37pp; English.  
 CC The invention relates to a method for treating pulmonary hypertension by  
 CC delivering a therapeutic composition, comprising an aerosolised antisense

CC endothelin-1 (ET-1) molecule, to the lungs of a patient. The composition  
 CC can be used in a method for determining the efficacy of the treatment for  
 CC e.g. when studying molecules and observing the effects of the composition  
 CC on an animal model system hypersensitive to antisense ET-1. The method is  
 CC useful for treating pulmonary hypertension. The aerosolised antisense  
 CC ET-1 molecule permits inhibition of the ET-1 transcription, which  
 CC relieves pulmonary hypertension. Its use avoids side effects caused by  
 CC alternative therapies. The present sequence represents the genomic DNA  
 CC sequence of human ET-1 gene.  
 SQ Sequence 12461 BP; 3372 A; 2728 C; 2816 G; 3544 T;

Query Match 6.9%; Score 17; DB 1; Length 12461;  
 Best Local Similarity 100.0%; Pred. No. 8;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 TCTGACTTGACAGCTC 54  
 DB 3610 TCTGACTTGACAGCTC 3594

RESULT 8  
 ID 037760 standard; DNA; 5100 BP.  
 AC 037760:  
 DT 28-JUN-1993 (first entry)  
 DE Glucosyltransferase I gene.  
 KW GT-1; Streptococcus; dental; caries; ss.  
 OS Streptococcus sobrinus.  
 FH key location/Qualifiers  
 FT cds 181..4959 /tag= a  
 FT J05023188-A.  
 PD 02-FEB-1993.  
 PF 25-JUL-1991: 186592.  
 PR 25-JUL-1991: JP-186592.  
 PA (FUKU/) FUKUI I.  
 DR WPI: 93-079449/10.  
 P-PSDB: R32925.  
 PT DNA sequence glucosyl:transferase-I - comprises Streptococcus  
 PT sobrinus DNA sequence with at least one nucleotide added or  
 PT deleted  
 PS Claim 3: Page 15; 29pp; Japanese.  
 CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtained by treating  
 CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with Sau3AI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and E. coli DH109  
 CC transformed with it. A GT-1-expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 SQ Sequence 5100 BP; 1575 A; 1051 C; 1111 G; 1363 T;

Query Match 6.5%; Score 16; DB 1; Length 5100;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 CTGAGGAAGCTTTT 205  
 DB 5004 CTGAGGAAGCTTTT 5019

RESULT 9  
 ID T19623 standard; cDNA to mRNA; 231 BP.  
 AC T19623:  
 DT 28-JUN-1996 (first entry)  
 DE Human gene signature HUMGS00691.  
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.

OS Homo sapiens.  
 PN W09514772-R1.  
 PD 01-JUN-1995.  
 PF 11-NOV-1994: J01916.  
 PR 12-NOV-1993: JP-355504.  
 PA (MATS/) MATSUBARA K.  
 PI (OKUBO/) OKUBO K.  
 PI Matsubara K, Okubo K;  
 DR WPI: 95-206931/27.  
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 PS Claim 1: Page 434; 2245pp; Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in T19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(I) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SQ Sequence 231 BP; 80 A; 29 C; 35 G; 80 T;

Query Match 6.5%; Score 16; DB 1; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 CTTTGTGTTAAACCA 215  
 DB 124 CTTTGTGTTAAACCA 139

RESULT 10  
 ID T15718 standard; cDNA to mRNA; 2139 BP.  
 AC T15718:  
 DT 18-JUL-1996 (first entry)  
 DE Human cDNA encoding gliocyte tumour cell T986-derived growth factor.  
 KW Growth factor; glial cell; glioma; neuron; haematopoietic; immune;  
 KW nervous system; tissue repair; ss.  
 OS Homo sapiens.  
 FH key location/Qualifiers  
 FT cds 192..1643 /tag= a  
 FT 192..1643 /transl\_except= pos:444..446, aa:Glu  
 FT 192..251 /transl\_except= pos:447..449, aa:Gly  
 FT 192..251 /transl\_except= pos:1224..1226, aa:Pro  
 FT mat\_peptide 252..1640 /tag= b  
 FT 252..1640 /tag= c  
 PN J08000271-A.  
 PD 09-JAN-1996.  
 PF 24-JUN-1994: 164860.  
 PR 24-JUN-1994: JP-164860.  
 PA (ONOX ) ONO PHARM CO LTD.  
 DR WPI: 96-091655/10.  
 P-PSDB: R90826.  
 PT DNA encoding polypeptide and related vectors, host cells and  
 PT antibodies - useful in treatment of incomplete or abnormal growth of  
 PT glial, neural or haematopoietic cells  
 PS Claim 1: Page 12-14; 15pp; Japanese.  
 CC T15718 encodes a human growth factor derived from the human gliocyte  
 CC tumour cell line T986. The growth factor can be used for the

CC treatment of diseases related to incomplete growth of glial, neuron  
CC or haematopoietic cells or diseases related to lowered or enhanced  
CC function of the immune and nervous systems. The growth factor is  
CC also useful as a tissue repair agent.  
SQ Sequence 2139 BP; 668 A; 378 C; 546 G; 547 T;

Query Match 6.5%; Score 16; DB 1; Length 2139;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 200 CTTTGTGTTAAACCA 215  
DB 2018 CTTTGTGTTAAACCA 2033

## RESULT 11

T18830  
ID T18830 standard; cDNA to mRNA; 2139 BP.  
AC T18830;

DE 18-JUL-1996 (first entry)  
DT Gliocyte tumour cell T98G-derived growth factor-hybridising cDNA.  
KW Growth factor; glial cell; glioma; neuron; haematopoietic; immune;  
KW nervous system; tissue repair; ss.  
OS Synthetic.  
PN J08000271.A.  
PD 09-JAN-1996.  
PE 24-JUN-1994; 164860.  
PF 24-JUN-1994; JP-164860.  
PA (ONOX ) ONO PHARM CO LTD.  
WP: 96-091655/10.

PT DNA encoding polypeptide and related vectors, host cells and  
PT antibodies - useful in treatment of incomplete or abnormal growth of  
PT glial, neural or haematopoietic cells  
PS Claim 5: Page 11-12; 15pp; Japanese.  
CC T15719 and T18830 are cDNA sequences capable of hybridising to DNA  
CC encoding a human growth factor, derived from the human gliocyte  
CC tumour cell line T98G. The growth factor can be used for the  
CC treatment of diseases related to incomplete growth of glial, neuron  
CC or haematopoietic cells or diseases related to lowered or enhanced  
CC function of the immune and nervous systems. The growth factor is  
CC also useful as a tissue repair agent.  
SQ Sequence 2139 BP; 668 A; 378 C; 546 G; 547 T;

Query Match 6.5%; Score 16; DB 1; Length 2139;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 200 CTTTGTGTTAAACCA 215  
DB 2018 CTTTGTGTTAAACCA 2033

## RESULT 12

T22284/C  
ID T22284 standard; cDNA to mRNA; 390 BP.  
AC T22284;

DE 13-SEP-1996 (first entry)  
DT Human gene signature HUMGS03856.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PE 11-NOV-1994; J01916.  
PF 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WP: 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1: Page 1090-91; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 390 BP; 71 A; 119 C; 96 G; 100 T;

Query Match 6.5%; Score 16; DB 1; Length 390;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 AAGCCACTGAGCA 196  
DB 248 AAGCCACTGAGCA 233

## RESULT 13

T23588  
ID T23588 standard; cDNA to mRNA; 287 BP.  
AC T23588;

DE 02-SEP-1996 (first entry)  
DT Human gene signature HUMGS05441.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PE 11-NOV-1994; J01916.  
PF 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WP: 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1: Page 1405; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 287 BP; 83 A; 56 C; 57 G; 89 T;

Query Match 6.5%; Score 16; DB 1; Length 287;

Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 94 TCATTGCAATAGTAC 109  
DB 188 TCATTGCAATAGTAC 203

RESULT 14  
ID T86242 standard; cDNA to mRNA; 1593 BP.  
AC T86242;  
BT 07-JAN-1998 (first entry)  
DE cDNA for birch pollen co-factor-independent phosphoglycerate mutase.  
KW Confactor-independent phosphoglycerate mutase; PGM-1; E.C. 5.4.21;  
KW birch; pollen; allergy; plant allergen; panallergen; B cell;  
KW T cell; epitope; immunotherapy; detection; diagnosis; hay fever;  
KW conserved; ds.  
OS Betula verrucosa.  
FH location/Qualifiers  
FT 1. 1593  
FT CDS  
FT /tag= a  
FT /note= "no start codon given"  
PN MO9705258-A2.  
PD 13-FEB-1997.  
PF 02-AUG-1996; AT0141.  
PR 02-AUG-1995; AT-001320.  
PA (BIOM-) BIOMACH PROD N. HANDEL'S GMBH.  
PI Breitenbach M. Ebner C. Engel E. Ferreira F. Jilek A.  
PI Kraft D. Richter K. Rheinberger H.  
DR WPI: 97-14595/13.  
DR P-PDB: W25501.  
PT New recombinant DNA encoding plant phosphoglycerate mutase or its  
PT antigenic epitope(s) - useful for diagnosis or treatment of  
PT allergies to pollen and plant-derived foods  
PS Claim 1: Fig 1; 160p; German.  
CC T86242 encodes birch pollen co-factor-independent phosphoglycerate  
CC mutase (PGM-1). PGM-1 is a highly conserved plant allergen (panallergen)  
CC which can cause cross-reactivity in patients allergic to pollen and  
CC plant-derived foods. PGM-1 and its B cell and T cell epitopes can be  
CC used for the in vitro detection of allergy against PGM-1, by measuring  
CC serum IgE or a cellular reaction. They can also be used in immunotherapy  
CC and will not cause an autoimmune response because PGM-1 is significantly  
CC different from the human enzyme, which is co-factor dependent.  
SQ Sequence 1593 BP; 424 A; 301 C; 433 G; 435 T;

Query Match 6.5%; Score 16; DB 1; Length 1593;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 227 TTTTGAATGACACA 242  
DB 279 TTTTGAATGACACA 294

RESULT 15  
ID V52215/c  
AC V52215 standard; DNA; 16995 BP.  
BT 23-OCT-1998 (first entry)  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:82.  
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.  
OS Streptococcus pneumoniae.  
PN WO9618931-A2.  
PD 07-MAY-1998.  
PF 30-OCT-1997; U19588.  
PR 31-OCT-1996; US-029960.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
PI Kunsch CA, Rosen CA;  
DR WPI: 96-27225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
PS Claim 1: Page 651-660; 1409pp; English.  
CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridize to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
SQ Sequence 16995 BP; 4992 A; 3235 C; 3669 G; 5099 T;

Query Match 6.5%; Score 16; DB 1; Length 16995;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 217 AAAACATTCTTTGA 232  
DB 586 AAAACATTCTTTGA 571

Search completed: August 5, 1999, 17:41:08  
Job time: 6303 sec

This Page Blank (uspto)

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:41:14 : Search time 3268.17 Seconds  
(without alignments)  
241.333 Million cell updates/sec

Title: US-09-049-696-14

Perfect score: 248  
Sequence: 1 ACCCTAGAGCGGAAATTCAC.....TTGAAATGCGACAGATCTT 248

Scoring table: OLIGO\_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl:.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl.\*  
8: gb\_pl2.\*  
9: gb\_pl1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_st.\*  
14: gb\_sts.\*  
15: gb\_sy.\*  
16: gb\_un.\*  
17: gb\_vl.\*  
18: em\_fun.\*  
19: em\_hcg.\*  
20: em\_hum1.\*  
21: em\_hum2.\*  
22: em\_in.\*  
23: em\_om.\*  
24: em\_or.\*  
25: em\_ov.\*  
26: em\_pat.\*  
27: em\_ph.\*  
28: em\_pl.\*  
29: em\_ro.\*  
30: em\_sts.\*  
31: em\_sy.\*  
32: em\_un.\*  
33: em\_vl.\*  
34: gb\_hcg1.\*  
35: gb\_hcg2.\*  
36: gb\_in1.\*  
37: gb\_in2.\*  
38: em\_ba1.\*  
39: em\_ba2.\*  
40: em\_hum3.\*  
41: em\_hum4.\*  
42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	248	100.0	878	5	195746	Sequence 8

2	248	100.0	3311	11	AF039400	AF039400 Homo sapi
3	248	100.0	2826	42	AF127036	AF127036 Homo sapi
4	169	68.1	35278	11	AF039401	AF039401 Homo sapi
5	19	7.7	90735	8	ATF28P10	AL049655 Arabidops
6	19	7.7	173766	11	AC004049	AC004049 Homo sapi
7	18	7.3	11719	7	SCDSPISA	X71622 Saccharomyc
8	18	7.3	5958	7	SCYKL033W	Z28033 S.cerevisia
9	18	7.3	81341	8	ATAC007213	AC007213 Arabidops
10	18	7.3	139190	9	HS929C8	AL049994 Human DNA
11	18	7.3	120385	10	HSJ437116	AL049570 Human DNA
12	18	7.3	173452	11	AC003013	AC003013 Human PAC
13	18	7.3	181098	11	AC004216	AC004216 Homo sapi
14	18	7.3	45893	12	RATINTELA	M73048 Rat alpha-1
15	18	7.3	188682	34	CEY47H9_4	AC005050 Homo sapi
16	18	7.3	82746	34	CEY47H9_4	Continuation (5 of
17	18	7.3	110000	34	CEY53C10_0	293340 Caenorhabdi
18	18	7.3	37384	36	CELE04F6	U28943 Caenorhabdi
19	18	7.3	29138	36	CELF52D1	AF026308 Caenorhab
20	18	7.3	75342	36	CEY47H9C	AL032657 Caenorhab
21	18	7.3	35840	37	CELC27H5	U14635 Caenorhabdi
22	17	6.9	8732	1	BACADDA	M63489 Bacillus su
23	17	6.9	210440	1	BSUB0006	Z99109 Bacillus su
24	17	6.9	53533	1	BSY09476	Y09476 B.subtilis
25	17	6.9	12631	2	AE001382	AE001282 Chlamydia
26	17	6.9	36783	2	AF078135	AF078135 Leptospir
27	17	6.9	2288	4	XELMTOCHN	L11231 Xenopus lae
28	17	6.9	3679	4	XLESCGAM2	Y12001 X.laevis mr
29	17	6.9	1308	5	AR005096	AR005096 Sequence
30	17	6.9	82415	7	AB005244	AB005244 Arabidops
31	17	6.9	87841	7	AB005247	AB005247 Arabidops
32	17	6.9	41087	7	AB024029	AB024029 Arabidops
33	17	6.9	78921	7	AB028621	AB028621 Arabidops
34	17	6.9	111767	8	AF058914	AF058914 Arabidops
35	17	6.9	79272	8	ATAC002338	AC002338 Arabidops
36	17	6.9	87080	8	ATAC004747	AC004747 Arabidops
37	17	6.9	104861	8	ATG93215	U93215 Arabidops
38	17	6.9	3803	9	AB018366	AB018366 Homo sapi
39	17	6.9	100000	9	AP000018	AP000018 Homo sapi
40	17	6.9	79482	9	AP000029	AP000029 Homo sapi
41	17	6.9	100000	9	AP000032	AP000032 Homo sapi
42	17	6.9	102456	9	AP000083	AP000083 Homo sapi
43	17	6.9	100000	9	AP000160	AP000160 Homo sapi
44	17	6.9	143747	9	HS370M22	Z82306 Human DNA s
45	17	6.9	186510	9	HS451B15	Z98050 Human DNA s

## ALIGNMENTS

RESULT	1	
LOCUS	195746	878 bp DNA
DEFINITION	Sequence 8 from patent US 5733748.	PAT
ACCESSION	195746	17-JUL-1998
NID	93940216	
VERSION	195746.1	GI:3940216
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unclassified.	
REFERENCE	1 (bases 1 to 878)	
AUTHORS	Yu, G. and Rosen, C.	
TITLE	Colon specific genes and proteins	
JOURNAL	Patent: US 5733748-A 8 31-MAR-1998;	
FEATURES	Location/Qualifiers	
source	1..878	
BASE COUNT	257 a 179 c 188 g 241 t	13 others
ORIGIN	/organism="unknown"	

Query Match 100.0%: Score 248; DB 5; Length 878;  
Best Local Similarity 100.0%: Pred. No. 6.2e-131;

Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTGAAGCGGAAATTCACGGGGGAGCTCTCATTAATCTGACTGGACAGCTCTGGG 60  
|||||  
DB 306 ACCTGAAGCGGAAATTCACGGGGGAGCTCTCATTAATCTGACTGGACAGCTCTGGG 365  
|||||  
QY 61 ATGATTATGACCTGAGAGAGCTCACAAGTATTCATTCGATAAAGTACAGTATCTTG 120  
|||||  
DB 366 ATGATTATGACCTGAGAGAGCTCACAAGTATTCATTCGATAAAGTACAGTATCTTG 425  
|||||  
QY 121 ATCTCAGAGACAGTTCATTAATCTCTTCAAGTGAATCTGCTGCTCATCCCAAG 180  
|||||  
DB 426 ATCTCAGAGACAGTTCATTAATCTCTTCAAGTGAATCTGCTGCTCATCCCAAG 485  
|||||  
QY 181 AAGCCACTCTGAGAGAGTCTTTTGTTTAAACCAAGAAACATTTGTAAGATGCA 240  
|||||  
DB 486 AAGCCACTCTGAGAGAGTCTTTTGTTTAAACCAAGAAACATTTGTAAGATGCA 545  
|||||  
QY 241 CAGATCTT 248  
|||||  
DB 546 CAGATCTT 553  
|||||

RESULT 2  
AF039400 3311 bp mRNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,  
DEFINITION complete cds.  
ACCESSION AF039400  
AF039400  
NID 94009457  
VERSION AF039400.1 GI:4009457  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreier,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
REFERENCE 2 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
SOURCE Location/Qualifiers  
1..3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue-type="small intestine"  
1..3311  
/gene="hCLCA1"  
352..3096  
/gene="hCLCA1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"

translation="MGPFKSSVFTLLIHLLEGALSNSLIQLNNGYEGIVVAIDPNP  
EDETLIQIDMTVAQLYLEATGRFRFKNVAILLPETWKTADYVRKLEFYKNA  
DVLVAESTPGNDPEYTEMONGCEGERHILPDLFAGKLAIEGPGKAFYENWA  
LRMGFVYNDDEKFEYLSNGRIQAVRCASGIGTNVYKKGSGCYTRKCTFKYNGI  
YKGCERFVLSROTERKASIMEFAHVDISVECTGNNKRAPKQNKCKLSTWEYI  
RDSDEPKTTPMTQPPNPFTSLQIQORIVCLVDKSGSMATGNRLNRLNAGQLEPL  
LQIVELGSWGVTFDSAAHVOSLIQINSGSDRLAKRLPAAASGSGTICGLRBA

FTVIRKRYPIDGSEIVLLTDGEDNTISGCENENKOSGAIHTVALGPSAOEIELSK  
MTGELQYASDPVOYNNGLIDAFGLSSGNASVRSIOLSKGLTLNOSMMNGTIV  
DSYGRDLFLITWTQPOLIMDPGSGOGGVYDKRMVYIOIPGIKAGTMY  
SLAASQTLITLVSRASNAIIPITVTSKNTDSEPSPIVYVYANIRGCAPIIYA  
SVTLISVYNGKYTLELDNGADATKRDGVYSRFTTYDNGRISVYRALGYN  
AARRVYIPQSGALYIPGWLENDIIONPREPEINKDDVYHKOVCESRISGGSFVAS  
DVPNAPIDLEFPQGITDLKAEIHGSLINLWTAPGDYDGHGAKYIIRISTSID  
LRDFNLSLOVNTALIPKEANSEVEFLFEPENTFENGTDLEFIAQAVDKLSKI  
SNIRVSLFIPQTPPETPSPDETSAPCPNINHINSTITPIGHILKIMKWIQELSLIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN  
Query Match 100.0%; Score 248; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 6, 6e-131;  
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTGAAGCGGAAATTCACGGGGGAGCTCTCATTAATCTGACTGGACAGCTCTGGG 60  
|||||  
DB 2624 ACCTGAAGCGGAAATTCACGGGGGAGCTCTCATTAATCTGACTGGACAGCTCTGGG 2683  
|||||  
QY 61 ATGATTATGACCTGAGAGAGCTCACAAGTATTCATTCGATAAAGTATCTTG 120  
|||||  
DB 2684 ATGATTATGACCTGAGAGAGCTCACAAGTATTCATTCGATAAAGTATCTTG 2743  
|||||  
QY 121 ATCTCAGAGACAGTTCATTAATCTCTTCAAGTGAATCTGCTGCTCATCCCAAG 180  
|||||  
DB 2744 ATCTCAGAGACAGTTCATTAATCTCTTCAAGTGAATCTGCTGCTCATCCCAAG 2803  
|||||  
QY 181 AAGCCACTCTGAGAGAGTCTTTTGTTTAAACCAAGAAACATTTGTAAGATGCA 240  
|||||  
DB 2804 AAGCCACTCTGAGAGAGTCTTTTGTTTAAACCAAGAAACATTTGTAAGATGCA 2863  
|||||  
QY 241 CAGATCTT 248  
|||||  
DB 2864 CAGATCTT 2871  
|||||

RESULT 3  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCI)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 2826)  
AUTHORS Agnel,M., Vermaut,T. and Culouscou,J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel,M. and Culouscou,J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des Carriers, Rueil-Malmaison 92500, France  
FEATURES  
SOURCE Location/Qualifiers  
1..2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue-type="small intestine; colon"  
1..2826  
/gene="CACCI"  
5..2749  
/gene="CACCI"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1



/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPFKSSVFILILHLEGLSNSLIQLNNGYEGIVVADPNV  
EDFTLQIKDMYQASLYLEATGRKRYFNVAIILPETKTKADYRRLPLEYKNA  
DLVYAEETPPRNDPEPTBOMNGCEKGRHILPDLFAGKLAIEYQGAFTYENAH  
LRMGVDEYINDEKFTLSNGRIQAVRCSAGITGVNKKCOGSCYTRCFNKVTGL  
YKCECEVLOSROTERKASIMFAOHVDSIVECTEONHKKAPNOKNCINRSTWEI  
RDSDFKFTPTPTOPNPFTSLQIGRIYCLVDKSGSMATGNRLNRLNQACFL  
LQIVELSGWGMVTFDSAAHVOSELIQINSQSDRLAKRLPAAAGSGTSIGLSRA  
FTVIRKRYPTDSEIYVLTDEGNTISGCENEVOSGAIHTVALGPSAAQELBELSK  
MTGLOTYASDVOVNNGLIDAFGALSNGAVSORSTQLESKGLTLQNSQWNGTIVY  
DSYVGKDTLFLITWTOPPOLILMDPSGOKGQGVNDNTMTALQIPGIAKGVTKWY  
SLQASQDTLITVTSRASNAIPLPIVTSKINKDTSKEPSPVAVANIRQASPIIRA  
SVTALIESVNGKIVTELLDNGAGADAKDGVYSRYETTDVQKQVCSRTSGSEFVAS  
AARRVYIPQOOGALYIPGWIENDEIOMNPPREINKNDYDQKQVCSRTSGSEFVAS  
DVNAPAPIPDLFPPGOITDLKAEIHGSLINLTMTAPGDYDHGTAHKYIIRISTIID  
LRQFNESLOYNTALIPKANSSEVFLPEPNTBNGTDLTAIOADVNDLAKSET  
SNIRVSLFIPQIPPTPPDETSAACPNIHINSTIPGHIILKIMKMWIGELQSLIA

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN  
Query Match 100.0%; Score 248; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 6.5e-111;  
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 ACCTAAGGCGGAATTCACGGGGGAGTCTCATTAATCTGACTTGACAGCTCCTGGGG 60  
Db 2277 ACCTAAGGCGGAATTCACGGGGGAGTCTCATTAATCTGACTTGACAGCTCCTGGGG 2336  
Oy 61 ATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAAATAGTACAGTATCTTG 120  
Db 2337 ATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAAATAGTACAGTATCTTG 2396  
Oy 121 ATCTAGAGACAGTTCATGATCTCTTCAAGTAACTCTCTGCTCATCCCAAGG 180  
Db 2397 ATCTAGAGACAGTTCATGATCTCTTCAAGTAACTCTCTGCTCATCCCAAGG 2456  
Oy 121 AAGCAACTCTGAGGAGCTTTTGTGTTAAACGAGAAACATCTTTTGAAGATGGCA 240  
Db 2457 AAGCAACTCTGAGGAGCTTTTGTGTTAAACGAGAAACATCTTTTGAAGATGGCA 2516  
Oy 241 CAGATCTT 248  
Db 2517 CAGATCTT 2524  
RESULT 4  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS AF039401 35278 bp DNA PRI 15-DEC-1998  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hclcal1) gene.  
ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human CLCA1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL GENOMICS 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.

TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1. .35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1. .1617  
/gene="hclcal1"  
1. .33522  
/gene="hclcal1"  
767. .1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587. .1596  
/gene="hclcal1"  
/note="putative"  
1618. .1708  
/gene="hclcal1"  
/number=1  
join(1618. .1708,1962. .2383,6667. .6807,6979. .7126,9696. .9801,15454. .15631,18591. .18809,19783. .20010,22253. .22427,24522. .24628,26644. .26859,27447. .27708,28766. .28936,31840. .32079,32919. .33322)  
/gene="hclcal1"  
/product="calcium-dependent chloride channel-1"  
1962. .2383  
/gene="hclcal1"  
/number=2  
join(2222. .2383,6667. .6807,6979. .7126,9696. .9801,15454. .15631,18591. .18809,19783. .20010,22253. .24522. .24628,26644. .26859,27447. .27708,28766. .28936,31840. .32079,32919. .33310)  
/gene="hclcal1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPFKSSVFILILHLEGLSNSLIQLNNGYEGIVVADPNV  
EDFTLQIKDMYQASLYLEATGRKRYFNVAIILPETKTKADYRRLPLEYKNA  
DLVYAEETPPRNDPEPTBOMNGCEKGRHILPDLFAGKLAIEYQGAFTYENAH  
LRMGVDEYINDEKFTLSNGRIQAVRCSAGITGVNKKCOGSCYTRCFNKVTGL  
YKCECEVLOSROTERKASIMFAOHVDSIVECTEONHKKAPNOKNCINRSTWEI  
RDSDFKFTPTPTOPNPFTSLQIGRIYCLVDKSGSMATGNRLNRLNQACFL  
LQIVELSGWGMVTFDSAAHVOSELIQINSQSDRLAKRLPAAAGSGTSIGLSRA  
FTVIRKRYPTDSEIYVLTDEGNTISGCENEVOSGAIHTVALGPSAAQELBELSK  
MTGLOTYASDVOVNNGLIDAFGALSNGAVSORSTQLESKGLTLQNSQWNGTIVY  
DSYVGKDTLFLITWTOPPOLILMDPSGOKGQGVNDNTMTALQIPGIAKGVTKWY  
SLQASQDTLITVTSRASNAIPLPIVTSKINKDTSKEPSPVAVANIRQASPIIRA  
SVTALIESVNGKIVTELLDNGAGADAKDGVYSRYETTDVQKQVCSRTSGSEFVAS  
AARRVYIPQOOGALYIPGWIENDEIOMNPPREINKNDYDQKQVCSRTSGSEFVAS  
DVNAPAPIPDLFPPGOITDLKAEIHGSLINLTMTAPGDYDHGTAHKYIIRISTIID  
LRQFNESLOYNTALIPKANSSEVFLPEPNTBNGTDLTAIOADVNDLAKSET  
SNIRVSLFIPQIPPTPPDETSAACPNIHINSTIPGHIILKIMKMWIGELQSLIA

BASE COUNT	ORIGIN
11245 a	6870 c 7230 g 9933 t
exon	19783. .20010 /gene="h1c1a1"
exon	/number=8 22253. 22427 /gene="h1c1a1"
exon	/number=9 24522. 24628 /gene="h1c1a1"
exon	/number=10 26644. 26859 /gene="h1c1a1"
exon	/number=11 27447. 27708 /gene="h1c1a1"
exon	/number=12 28766. 28936 /gene="h1c1a1"
exon	/number=13 31840. 32079 /gene="h1c1a1"
exon	/number=14 32919. 33522 /gene="h1c1a1"
exon	/number=15

	Query Match	68.1%;	Score 169;	DB 11,	Length 35278;	
	Best Local Similarity	100.0%;	Pred. No. 8e-86;			
	Matches 169;	Conservative	0;	Mismatches	0;	Indels
						Gaps 0;
OY	80	AGCTCACAAGTATATCATTCGATAAGTACAGATATTCTTGATCTCAGAGACAAAGTTCAA	139			
Db	32917	AGCTCACAGAAGTATATCATTCGATAAGTACAGATATTCTTGATCTCAGAGACAAAGTTCAA	32976			
OY	140	TGAATCTCTTCAAGGAAACTACAGCTCGCTCCATCCCAAGGAAAGGCCAATCTGAGGAAGT	199			
Db	32977	TGAATCTCTTCAAGGAAACTACAGCTCGCTCCATCCCAAGGAAAGGCCAATCTGAGGAAGT	33036			
OY	200	CTTTTTGTTTAAACCAGAAAACATTACTTTTGGAAAAATGGCAGACAGATCTT	248			
Db	33037	CTTTTTGTTTAAACCAGAAAACATTACTTTTGGAAAAATGGCAGACAGATCTT	33085			
RESULT	5					
LOCUS	ATF28P10/c					
DEFINITION	Arabidopsis thaliana DNA chromosome 3, BAC clone F28P10.					
ACCESSION	AL049655					
NID	94678291					
VERSION	AL049655.1					
KEYWORDS	GI:4678291					
SOURCE	thale cress.					
ORGANISM	Arabidopsis thaliana					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
	euphyllophytes; Spermatophytes; Magnoliophyta; eudicotyledons; core					
	eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;					
	Arabidopsids.					
REFERENCE	1 (bases 1 to 90735)					
AUTHORS	Choinsne,N., Robert,C., Brottier,P., Wincker,P., Catolico,L.,					
	Artiguenave,F., Saurin,M., Weissenbach,J., Meyers,H.W.,					
	Mayer,K.F.X., Lemcke,K., Scheller,C., Quetier,F. and Salanoubat,M.					
	Unpublished					
JOURNAL	2 (bases 1 to 90735)					
REFERENCE	EU Arabidopsis sequencing project.					
AUTHORS	Direct Submission					
TITLE	Submitted (09-JUN-1999) MIPS, at the Max-planck-Institut fuer					
JOURNAL	Biochemie, Am Klopferstritz 18a, D-82152 Martinsried, FRG, E-mail:					
	schnell@lewis.biochem.mpg.de,mayer@mips.biochem.mpg.de Project					
	Coordinator: Marcel Salanoubat and Francis Quetier, Groupement					
	d'Interet Public, Centre National de Sequençage - GENOSCOPE; 2 rue					
	Gaston Cremieux, BP191, 91006 Evry Cedex, France;					

COMMENT	FEATURES	SOURCE
<a href="http://www.genoscope.cns.fr">http://www.genoscope.cns.fr</a> Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a>	Location/Qualifiers 1 90735	

gene  
2..462  
/gene="F28P10.10"  
complement(2..72)  
/gene="F28P10.10"  
/number=1  
complement(join(>..72,159..462))  
/gene="F28P10.10"  
/EC.number="6.3.3.1"  
/codon.start=1  
/product="phosphoribosylformylglycinamide cyclo-ligase  
(fragment)"  
/protein\_id="CAB41083.1"  
/db\_xref="PID:e1425781"  
/db\_xref="PID:g4678292"  
/db\_xref="GI:4678292"  
/translation="MEARIIQSSSCSYLLVAVNSRRSSVSPPEVSQAOTTRR  
TRVLSMSKRGKRTDKDDIDSLNYSKSGVDIDAGELVKRIAKMAPGIGGGGFLPG  
DSYLVAGTIDGGTKLKLAPEFGI"  
complement(73..158)  
/gene="F28P10.10"  
/number=1  
complement(159..462)  
/gene="F28P10.10"  
/number=2  
complement(1087..1132)  
/gene="F28P10.20"  
/number=1  
1087..5668  
/gene="F28P10.20"  
complement(join(1087..1132,1218..1388,1475..1599,  
2041..2153,3063..3153,3841..3963,4212..4336,4837..4959,  
5582..5668))  
/gene="F28P10.20"  
/codon.start=1  
/product="hypothetical protein"  
/protein\_id="CAB41084.1"  
/db\_xref="PID:e1425782"  
/db\_xref="PID:g4678293"  
/db\_xref="GI:4678293"  
/translation="MDYITREKMDKTLVTRLEKGYLAKIRPKDSKSELRDFNSN  
NGEFLANSGSPGLDLVDLEGLFKESMTQGGSSSRDSESSSSLESNNPRRS  
ASDLPQRRKDEFNMRGNODTHEVTRPASAALNQLDRTKRTLTSMTMDYTR  
EMMDIKLVTRLEKGYLAKIRKDSKSELRDFSINNGEELNNEDESRLLDVL  
GLKFEEMTQVGSGSRSESESSLSLTPNPRRSASDSLPHORRSVASQSGA  
ATLSYRKDESNMRYTDEDPDEYMRASTALENLQDKRTLTSMKRVKDGITSEEE  
GKD"  
complement(1133..1217)  
/gene="F28P10.20"  
/number=1  
complement(1218..1388)  
/gene="F28P10.20"  
/number=2  
complement(1389..1474)  
/gene="F28P10.20"  
/number=2  
complement(1475..1599)  
/gene="F28P10.20"  
/number=3  
complement(1600..2040)  
/gene="F28P10.20"  
/number=3  
complement(2041..2163)  
/gene="F28P10.20"

introns /number=4  
complement(2164..3062)  
/gene="F28P10.20"  
/number=4  
complement(3063..3153)  
/gene="F28P10.20"  
/number=5  
complement(3154..3840)  
/gene="F28P10.20"  
/number=5  
complement(3841..3963)  
/gene="F28P10.20"  
/number=6  
complement(3964..4211)  
/gene="F28P10.20"  
/number=6  
complement(4212..4336)  
/gene="F28P10.20"  
/number=7  
complement(4337..4836)  
/gene="F28P10.20"  
/number=7  
complement(4837..4959)  
/gene="F28P10.20"  
/number=8  
complement(4960..5581)  
/gene="F28P10.20"  
/number=8  
complement(5582..5668)  
/gene="F28P10.20"  
/number=9  
10758..12244  
/gene="F28P10.30"  
join(10758..11176,11291..11316,11668..11698,11844..11931,  
12065..12244)  
/gene="F28P10.30"  
/note="strong similarity to APERAL2, Arabidopsis  
thaliana, U12546"  
/codon\_start=1  
/product="APERAL2-like protein"  
/protein\_id="CAB41085.1"  
/db\_xref="PID:e1425783"  
/db\_xref="PID:g4678294"  
/db\_xref="GI:4678294"  
/translation="MLDLNLKIFSSYNEDQDKRVPLMISTGGEESNSSSSSTTDSAA  
RDAFLAFGLIKRDDLVPPPPPKHETGDLFPVADARNTIEFSEDSHMTLSLQ  
RNTQKMKSRGSRSSOYGVTFYRTGFWESHWDGKQVYLGFEDYAAARA  
YDRAIKFRGLADINFPVDDYRDIDMKMNKNKEFVQTLARESASFGROSSKTKGL  
ALQKCTQKTHDQHLFQNRHFFVHVF"  
exon 10758..11176  
/gene="F28P10.30"  
introns 11177..11290  
/gene="F28P10.30"  
exon 11291..11316  
/gene="F28P10.30"  
introns 11317..11667  
/gene="F28P10.30"  
exon 11668..11698  
/gene="F28P10.30"  
introns 11699..11843  
/gene="F28P10.30"  
exon 11844..11931  
/gene="F28P10.30"  
introns 11932..12064  
/gene="F28P10.30"  
exon 12065..12244

exon 12065..12244  
/gene="F28P10.30"  
exon complement(14432..16987)  
/gene="F28P10.40"  
/number=1  
14437..16987  
/gene="F28P10.40"  
complement(14432..16987)  
/gene="F28P10.40"  
/note="similarity to CRP1, Zea mays, PIR2:T01685; contains  
EST gb:H36636"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB41086.1"  
/db\_xref="PID:e1425784"  
/db\_xref="PID:g4678295"  
/db\_xref="GI:4678295"  
/translation="NRSLVFRKIPSRIRLNLRNNKRFCSQSOPFKSENPSPQRL  
LVGTSSENPVTSKVSILSKPEOKDASVYIDLNRNNPEALRFYNARFWRGS  
FEDGVEFWVLHILVSSPEYGRASDLIRVTSNPNPASVLSKLVDSAKSFEF  
VNSRAFNILNAYSKDQTDHADVIVNMLDVIPEFPYVNRILSALQKNSLLEAK  
ELYSRAVNIAGVDQNVTTOLMRASLREKPAEALVSRALERGAEPDSLLSYAVO  
ACCTIDLMANANSLIREMKERKICVPSQETSVTLASVKGNNMDATRLDEMLDSG  
ISANVVAATSLITGHCKRNDLVSAVLEFDKREKGRSPNSVTFSLIEMFRKNGEMK  
ALEFKKMEVIGLTPSVFHVHTIIQGLKQKHEBALNLEDFSEFTGLANFVGNLIL  
SMLKQKGTDEATELISMEBSRGIPNVASYNWMLGCRKQKNDLARIIVSNLEKG  
LKPNTYYSILIDCGFRNHDEQNALLEVNHMTSSNIEVGVYVOTIINGLCKVGQTSK

Query Match 7.7%; Score 19; DB 8; Length 90735;  
Best Local Similarity 100.0%; Pred. No. 2.8;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 AAACATTACTTTTGAAA 235  
DB 36248 AAACATTACTTTTGAAA 36230  
|||||  
AC004049/c 173766 bp DNA PRI 31-JUL-1998  
LOCUS Homo sapiens chromosome 4 clone B203023 map 4q25, complete  
DEFINITION  
ACCESSION AC004049  
VERSION 93366580  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 173766)  
Stone,N.E., Keseler,I.M., Schmutz,J.J., Shang,J., Cox,D.R. and  
Myers,R.M.  
TITLE Direct Submission  
JOURNAL Unpublished  
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1998) Department of Genetics, Stanford Human  
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA  
3 (bases 1 to 173766)  
Stone,N.E., Keseler,I.M., Schmutz,J.J., Shang,J., Cox,D.R. and  
Myers,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (31-JUL-1998) Department of Genetics, Stanford Human  
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA  
On Jul 31, 1998 this sequence version replaced gi:3088603.  
COMMENT Estimated Total Number of Errors is 0.8.  
Single Stranded 56144-56290  
STS Content:  
SHGC-50969 G33968



EKLAVEDDSLSPKADYVYVPIASLKPALGESOTEVILLIIFHLATCCSSNGK  
 SSQLQOPLPLTFLVSSKDNOKLITRDERKIGCVLHOFVSVNRQKHKRFSS  
 NKRPNLIPALGHSVTLKILQSPQNNLOKALASLEVLQDIIIDEMLSFLPG  
 NVSVFAKILTKRQIHYKVCVLTLEVLAKLVLYVDPDLIKVAKLIDELNSID  
 LKHEINOSFMNGPIVILTRDGTHTDTSMLATSGOINIALEAFIPKLKRNNSID  
 EALATFVSLITRCNSLNCKVAVSTVHLERDMSLPSHIVKIKRVVMDLAKL  
 SPITRFENADRLLSSISFATITIEKKNERPTMNEVRCRPFESLNSIEPSPINKER  
 ITQSSOLTYVNFENLESTNALIALPRISDMSLKKFTYHMSGLLERHINDVY  
 TELISEOVSPRQKIVALMTSTFKAMEKOPKEEVYLOESANYSSEVEVCL  
 IYLEFCNELSDQISMEIECKGIRKSEFAVCVLSIEFICAVMREEPQLIDYIT  
 VVDALASPEALIRYVOSCALRIADLYGSIIPNMLISVADYLVESISSRLNGMTER  
 VSOILMVICOLAGYEIENFKVIEIFELLDYHGYSDLCIOFOLFEIITLKKKK  
 YINDENMILKINONHISOSTFSPKMTDQOVLINLIDKTYQKDDITDNDVFLAKD  
 NPSNFQETFDKRLBPDDDEEKEEVESSKEYTDQWTSPIPSDYLKILQILG  
 YGERLTHPSKRLRVOILVRLIFPLSTOHNLIREVAVWSLSIIOCVLCSDSIV  
 OPACSCVEQMIRYSGDFVAKRFIELMOKLQCSFIIKEKIDIPVYHNEKKSISKRVK  
 PPPTENALVSVHMYHLEGVKITELYISEAVLEQIITCYQVVPVEKISSMLYVDI  
 VVKIRININ"

tRNA  
 5574..5647  
 /gene="ty(GTT)KIL - systematic name"  
 /note="tRNA-Valia - common name; anticodon gene: AAC"  
 5574..5647  
 /gene="ty(GTT)KIL - systematic name"  
 1983 a 952 c 1162 g 1861 t

BASE COUNT  
 ORIGIN

Query Match  
 Best Local Similarity 100.0%; Score 18; DB 7; Length 5958;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 AAGTCAATGATCTCTT 149  
 ||||||||||||||||  
 Db 4166 AAGTCAATGATCTCTT 4149

RESULT 9  
 ATRC007213/c ATAC007213 81341 bp DNA PLN 04-MAY-1999  
 LOCUS Arabidopsis thaliana chromosome II BAC T23015 genomic sequence,  
 DEFINITION complete sequence.  
 ACCESSION AC007213  
 NID 94689466  
 VERSION AC007213.4 GI:4689466  
 KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsids.  
 REFERENCE 1 (bases 1 to 81341)  
 AUTHORS Lin.X., Kaul.S., Shea.T.P., Fujii.C.Y., Shen.M., Vanaken.S.E.,  
 Barnstead.M.E., Mason.T.M., Bowman.C.L., Ronning.C.M., Beilto.M.,  
 Frazer.C.M. and Venter.J.C.  
 TITLE Arabidopsis thaliana chromosome II BAC T23015 genomic sequence  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 81341)  
 AUTHORS Lin.X. and Kaul.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-APR-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA, kline@tigr.org  
 REFERENCE 3 (bases 1 to 81341)  
 AUTHORS Lin.X.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-APR-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 REFERENCE 4 (bases 1 to 81341)  
 AUTHORS Lin.X.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-MAY-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA

## COMMENT

On Apr 27, 1999 this sequence version replaced gi:4580373.  
 Address all correspondence to:  
 Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr.  
 Rockville, MD 20850, USA  
 e-mail: kline@tigr.org  
 BAC clone T23015 is from Arabidopsis chromosome II and is near the  
 molecular marker RNS1.  
 The orientation of the sequence is from SP6 to T7 end of the BAC  
 clone.

Genes were identified by a combination of three methods: Gene  
 prediction programs including GRAL (available by anonymous ftp  
 from [atthir.ebm.ornl.gov](http://atthir.ebm.ornl.gov)), GeneFinder (Phil Green, University of  
 Washington), Genscan (Chris Burge,  
<http://genomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene  
 (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>). Searches of the  
 complete sequence against a peptide database and the Arabidopsis  
 EST database at TIGR (<http://www.tigr.org/tdb/at.html>).  
 Annotated genes are named to indicate the level of evidence for  
 their annotation. Genes with similarity to other proteins are named  
 after the database hits. Genes without significant peptide  
 similarity but with EST similarity are named as 'unknown' proteins.  
 Genes without protein or EST similarity that are predicted by more  
 than two gene prediction programs over most of their length are  
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
 predicted by tRNAscan-SE (Sean Eddy,  
<http://genome.mustl.edu/eddy/tRNAscan-SE/>). Simple repeats are  
 identified by repeatmasker (Arian Smit,  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of  
 genomic sequence that are not annotated as genes but have predicted  
 exons by GRAL are annotated as misc features.

## FEATURES

source  
 1..81341  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="II"  
 /map="RNS1"  
 /clone="T23015"  
 1..22779  
 /note="overlap with BAC clone T103  
 (AC006951:77026..99804)."  
 complement(2..53)  
 /note="exon predicted by xgrail, quality  
 marginal\_shadowexon"  
 47..105  
 /note="exon predicted by xgrail, quality marginal"  
 join(<1096..1170,1281..>1454)  
 /gene="T23015.1"  
 <1096..>1454  
 /gene="T23015.1"  
 /note="T23015.1"  
 /note="T103.22; predicted by genscan"  
 join(1096..1170,1281..1454)  
 /gene="T23015.1"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAD27903.1"  
 /db\_xref="PID:94689467"  
 /db\_xref="GI:4689467"  
 /translation="MKFAVKIMKVEDVEDEPFVKKKFRILRLDSKRRDPNRIPI  
 AFINSEKFPFLQLQFADNRNIRFRWGLSLTKNPKR"  
 1121..1179  
 /rpt\_family="(GAA)n"  
 /rpt\_start=1  
 /rpt\_end=1179  
 /rpt\_family="(TA)n"  
 complement(1891..11969)  
 join(<3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)  
 /gene="T23015.2"  
 <3068..>5572  
 /gene="T23015.2"  
 /note="T103.23; predicted by genscan"  
 join(3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)

misc\_feature  
 1..22779  
 /note="overlap with BAC clone T103  
 (AC006951:77026..99804)."  
 complement(2..53)  
 /note="exon predicted by xgrail, quality  
 marginal\_shadowexon"  
 47..105  
 /note="exon predicted by xgrail, quality marginal"  
 join(<1096..1170,1281..>1454)  
 /gene="T23015.1"  
 <1096..>1454  
 /gene="T23015.1"  
 /note="T23015.1"  
 /note="T103.22; predicted by genscan"  
 join(1096..1170,1281..1454)  
 /gene="T23015.1"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAD27903.1"  
 /db\_xref="PID:94689467"  
 /db\_xref="GI:4689467"  
 /translation="MKFAVKIMKVEDVEDEPFVKKKFRILRLDSKRRDPNRIPI  
 AFINSEKFPFLQLQFADNRNIRFRWGLSLTKNPKR"  
 1121..1179  
 /rpt\_family="(GAA)n"  
 /rpt\_start=1  
 /rpt\_end=1179  
 /rpt\_family="(TA)n"  
 complement(1891..11969)  
 join(<3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)  
 /gene="T23015.2"  
 <3068..>5572  
 /gene="T23015.2"  
 /note="T103.23; predicted by genscan"  
 join(3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)

repeat\_region  
 1121..1179  
 /rpt\_family="(GAA)n"  
 /rpt\_start=1  
 /rpt\_end=1179  
 /rpt\_family="(TA)n"  
 complement(1891..11969)  
 join(<3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)  
 /gene="T23015.2"  
 <3068..>5572  
 /gene="T23015.2"  
 /note="T103.23; predicted by genscan"  
 join(3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)

mRNA  
 1121..1179  
 /rpt\_family="(GAA)n"  
 /rpt\_start=1  
 /rpt\_end=1179  
 /rpt\_family="(TA)n"  
 complement(1891..11969)  
 join(<3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)  
 /gene="T23015.2"  
 <3068..>5572  
 /gene="T23015.2"  
 /note="T103.23; predicted by genscan"  
 join(3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)

gene  
 1121..1179  
 /rpt\_family="(GAA)n"  
 /rpt\_start=1  
 /rpt\_end=1179  
 /rpt\_family="(TA)n"  
 complement(1891..11969)  
 join(<3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)  
 /gene="T23015.2"  
 <3068..>5572  
 /gene="T23015.2"  
 /note="T103.23; predicted by genscan"  
 join(3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)

CDS  
 1121..1179  
 /rpt\_family="(GAA)n"  
 /rpt\_start=1  
 /rpt\_end=1179  
 /rpt\_family="(TA)n"  
 complement(1891..11969)  
 join(<3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)  
 /gene="T23015.2"  
 <3068..>5572  
 /gene="T23015.2"  
 /note="T103.23; predicted by genscan"  
 join(3068..3140,4009..4153,4330..4478,4600..4642,  
 4823..4982,5106..5204,5271..5353,5437..5572)

4823..4982,5106..5204,5271..5353,5437..5572)  
/gene="T23015.2"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="AAD27904.1"  
/db\_xref="pid:g4689468"  
/db\_xref="gi:4689468"  
/translation="MDRLMEATPKLESKIRVQNYVIFQAGHNVRIRASDPNPKRSQ  
NIRIGFYTRPKERSVSDPNRTKCVQOPRIALFPYERKVFTHICKSSNNPEERK  
TQIDGRDMSISILFALMDIRLVLFKAEFESEKTPAMPEVVSFPGYVALPYF  
ALMNPSPVSETELROMPLANLYSEKTAGVTLVGLGIILSYVGNAGDMTEFYOF  
RESKRIHTSLDFCLLSAFAPFMYNDMTTRKMFDSKMLPVSIVPLGSLYILLR  
PAVETIAPKDTASDPNO"  
5783..5841  
/note="exon predicted by xgrall, quality  
marginal\_shadowexon"  
complement(join(<6039..6203,6288..6494,6589..6723,  
6797..6928,7009..7126,7293..7522,7605..7778,7874..8005,  
8098..8370,8459..8539,8616..9142,9287..9366))  
/gene="T23015.3"  
complement(<6039..9366)  
/gene="T23015.3"  
/note="T103.24"  
complement(join(6039..6203,6288..6494,6589..6723,  
6797..6928,7009..7126,7293..7522,7605..7778,7874..8005,  
8098..8370,8459..8539,8616..9131))  
/gene="T23015.3"  
/codon\_start=1  
/product="putative acyl-CoA synthetase"  
/protein\_id="AAD27905.1"  
/db\_xref="pid:g4689469"  
/db\_xref="gi:4689469"  
/translation="MEDSGVNPMDSPKSGSPFYGIIGGIVALLVPLVILVNLGT  
KKGKRGVIVGSEGTMRHARPELVDPWGAAMPALFEOSCKKYSKDLDT  
REFIDKEITASDGRKEKHLGEGWQSYDEVERCNFASGVNNGHNDVAIF  
SDTRAEVIAFOGCFROSLITVITIASGEALLYSINETSILCDSKLTSAL  
QSLKTVNIIYIEEDGVNADSVNSGDTISISIVELGQKNVOPILPEKNE  
AVIMETSGTGLPKGVMTTGNLVATAGVAKVVPKIDKNTYATYALPIAHPELE  
IVFTSGAIGYSAMLTIDPSNKKYKTKDVALEPTITANTPALIDRYREGLVK  
VEKGGMAKTLFDTRIKRRLAIVDSGSPFAGLMDALVFKIRAVLGHIREML  
VGGAPLSFDSQRFINICWSPIGQYGLTECAATFSEMDPVGVRGPLPCGYVK  
LVSEDEGRISDPRGCEIIVGNSVTAGYFNNOKTIDEVYVADKGRMEYTGDI  
GREFHDCLEIVIDRKKDIVKLQHGELYSLGVEALSSNNVDINHADINSYCA  
LVPSRGALKEKMAEAGVHSEFALEKCAVKEVOOSLTKAKAKLEKLELPARI  
KLSEPTPESGITYALTKIRBOIKSKFDELSEKLA"  
complement(9369..9611)  
/note="exon predicted by xgrall, quality good"  
9776..9823  
/note="exon predicted by xgrall, quality  
marginal\_shadowexon"  
9873..9936  
/note="exon predicted by xgrall, quality  
marginal\_shadowexon"  
9984..10062  
/note="exon predicted by xgrall, quality  
excellent\_shadowexon"  
complement(10019..10097)  
/note="exon predicted by xgrall, quality excellent"  
complement(10162..10383)  
/note="exon predicted by xgrall, quality excellent"  
10185..10328  
/note="exon predicted by xgrall, quality  
marginal\_shadowexon"  
complement(10396..10430)  
/note="exon predicted by xgrall, quality good"  
10501..10610  
/note="exon predicted by xgrall, quality  
marginal\_shadowexon"  
complement(10558..10806)  
/note="exon predicted by xgrall, quality excellent"  
11012..11169  
/note="exon predicted by xgrall, quality  
marginal\_shadowexon"

misc-feature complement(11070..11174)  
/note="exon predicted by xgrall, quality excellent"  
join(<12075..12501,12559..12794,13308..13588,13712..13911,  
14130..14284,14372..14767)  
/gene="T23015.4"  
/note="T23015.4"  
/db\_xref="pid:g4689470"  
/db\_xref="gi:4689470"  
/translation="MPKTRGGGGRKREIAPRTKEVKAPEKVEKPAKAK  
VQAPAKREKQAPAKKAKVAPAKTRIACTIATATATATATATATATATATAT  
ESPMDDSTYDALKHPAETQENNOTDEVEDENKEBEKETETNELACANPVEEA

Query Match 7.3%; Score 18; DB 8; Length 81341;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 162 ACTGCTCATCCCAAG 179  
Db 44760 ACTGCTCATCCCAAG 44743  
|||||

RESULT 10  
HS929C8  
LOCUS  
DEFINITION  
ACCESSION  
MID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 139190)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (03-AUG-1998) E-mail enquiries: humquersanger.ac.uk  
Clone requests: clonerequestsanger.ac.uk  
On Jul 30, 1998 this sequence version replaced g1:3319855.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence is the entire insert of clone 929C8. The true right  
end of clone d11172A22 is at 36176 in this sequence. This sequence  
has been finished according to sequence map criteria as follows. An  
attempt is made to resolve all sequencing problems, such as  
compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome 22, constructed by the Sanger Centre Chromosome 22  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr22  
929C8 is from the human BAC library described in U-J. Kim et al.  
(1996) Genomics 34, 213-218. VECTOR: pBeloBAC11.  
Location/Qualifiers  
1..139190  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"

```
/clone="929c8"  
/map="q12.1-12.3"  
9. .649  
misc_feature  
/note="match: GSS B14128 clone A-926A6"  
437. .439  
variation  
/note="clone 1172A22; aga in this entry; substitution"  
/replace="aaa"  
1117. .1119  
variation  
/note="clone 1172A22; cac in this entry; substitution"  
/replace="cgc"  
1639. .1936  
repeat_region  
/note="AluY repeat: matches 1. .299 of consensus"  
1663. .1665  
variation  
/note="clone 1172A22; gta in this entry; substitution"  
/replace="gca"  
1882. .1884  
variation  
/note="clone 1172A22; gcc in this entry; substitution"  
/replace="gtc"  
1977. .1979  
variation  
/note="clone 1172A22; aca in this entry; substitution"  
/replace="ata"  
2192. .2194  
variation  
/note="clone 1172A22; ccg in this entry; substitution"  
/replace="ctg"  
complement(2746. .2852)  
repeat_region  
/note="MIR repeat: matches 180. .66 of consensus"  
3008. .3010  
variation  
/note="clone 1172A22; atg in this entry; insertion"  
/replace="ag"  
3128. .3130  
variation  
/note="clone 1172A22; ttc in this entry; substitution"  
/replace="tcc"  
3488. .3490  
variation  
/note="clone 1172A22; cat in this entry; substitution"  
/replace="cgt"  
3702. .3705  
variation  
/note="clone 1172A22; taga in this entry; substitution"  
/replace="tcta"  
3775. .3919  
repeat_region  
/note="MIR repeat: matches 37. .183 of consensus"  
3982. .4213  
repeat_region  
/note="MIR repeat: matches 33. .262 of consensus"  
complement(4295. .4343)  
repeat_region  
/note="MIR repeat: matches 150. .102 of consensus"  
4810. .4812  
variation  
/note="clone 1172A22; tcc in this entry; substitution"  
/replace="ttc"  
4923. .4946  
repeat_region  
/note="12 copies of CA 100% conserved; differs from  
552408"  
5077. .5245  
repeat_region  
/note="MIR repeat: matches 53. .215 of consensus"  
5900. .5933  
repeat_region  
/note="17 copies 2 mer ta 91% conserved"  
6251. .6253  
variation  
/note="clone 1172A22; ggg in this entry; substitution"  
/replace="gtg"  
6893. .6895  
variation  
/note="clone 1172A22; agg in this entry; substitution"  
/replace="acg"  
6894  
unSURE  
complement(7394. .7480)  
repeat_region  
/note="MIR2 repeat: matches 145. .56 of consensus"  
8618. .8817  
repeat_region  
/note="MIR repeat: matches 47. .260 of consensus"  
9045. .9340  
repeat_region  
/note="AluY repeat: matches 1. .301 of consensus"  
10153. .10180  
repeat_region  
/note="7 copies 4 mer gttt 100% conserved"  
10806. .11006  
repeat_region  
/note="L1MB5 repeat: matches 469. .672 of consensus"  
11192. .11241  
repeat_region  
/note="2 copies 25 mer 100% conserved"  
  
repeat_region  
11248. .11326  
/note="L1MB3 repeat: matches 673. .756 of consensus"  
complement(12026. .12102)  
repeat_region  
/note="MIR repeat: matches 139. .60 of consensus"  
12168. .12199  
repeat_region  
/note="8 copies 4 mer ttaa 91% conserved"  
12885. .13074  
repeat_region  
/note="MIR repeat: matches 28. .262 of consensus"  
complement(13043. .13222)  
repeat_region  
/note="MIR repeat: matches 213. .47 of consensus"  
13444. .13553  
repeat_region  
/note="MIR repeat: matches 23. .142 of consensus"  
13684. .13742  
repeat_region  
/note="MIR repeat: matches 99. .157 of consensus"  
13782. .13968  
repeat_region  
/note="MIR repeat: matches 20. .238 of consensus"  
complement(14018. .14310)  
repeat_region  
/note="AluX repeat: matches 293. .1 of consensus"  
15056. .15058  
variation  
/note="clone 1172A22; atg in this entry; substitution"  
/replace="acg"  
complement(15807. .16105)  
repeat_region  
/note="match: STS G03708"  
16508. .16532  
repeat_region  
/note="MIR repeat: matches 120. .144 of consensus"  
complement(16561. .16662)  
misc_feature  
/note="match: 5' H5237 genomic clone C22_219"  
complement(17048. .17099)  
repeat_region  
/note="MIR repeat: matches 147. .96 of consensus"  
complement(17374. .17419)  
repeat_region  
/note="MIR2 repeat: matches 139. .94 of consensus"  
complement(17940. .17995)  
repeat_region  
/note="MIR2 repeat: matches 138. .79 of consensus"  
17996. .18089  
misc_feature  
/note="match: 5' H5208 genomic clone C22_186"  
18577. .18770  
repeat_region  
/note="MIR repeat: matches 23. .218 of consensus"  
18835. .18837  
variation  
/note="clone 1172A22; ccc in this entry; substitution"  
/replace="ctc"  
18908. .18973  
repeat_region  
/note="MIR2 repeat: matches 75. .140 of consensus"  
18980. .19032  
repeat_region  
/note="MIR repeat: matches 86. .139 of consensus"  
19054. .19089  
repeat_region  
/note="9 copies 4 mer catt 89% conserved"  
19208. .19210  
variation  
/note="clone 1172A22; agg in this entry; substitution"  
/replace="aag"  
19783. .19785  
variation  
/note="clone 1172A22; aac in this entry; substitution"  
/replace="aac"  
19784. .19963  
repeat_region  
/note="MIR repeat: matches 43. .224 of consensus"  
19957. .19959  
variation  
/note="clone 1172A22; agt in this entry; substitution"  
/replace="aat"  
20163. .20165  
variation  
/note="clone 1172A22; gta in this entry; substitution"  
/replace="gca"  
20596. .20694  
repeat_region  
/note="MIR repeat: matches 164. .262 of consensus"  
20789. .20791  
variation  
/note="clone 1172A22; ccg in this entry; substitution"  
/replace="cag"  
21249. .21251  
variation  
/note="clone 1172A22; ccg in this entry; substitution"  
/replace="ctg"  
21433. .21435  
variation  
/note="clone 1172A22; ggg in this entry; substitution"  
/replace="gag"  
21761. .21763  
variation  
/note="clone 1172A22; ggg in this entry; substitution"
```

```

variation      /replace="gtg"
                22059. .22060
                /note="clone 1172A22; tg in this entry; deletion"
variation      /replace="tgggg"
                22298. .22300
                /note="clone 1172A22; gcc in this entry; substitution"
repeat_region  /replace="gtc"
                22563. .22676
                /note="MIR repeat: matches 73. .191 of consensus"
repeat_region  22993. .23070
                /note="MIR repeat: matches 70. .146 of consensus"
repeat_region  23422. .23512

Query Match      7.3%; Score 18; DB 9; Length 139190;
Best Local Similarity 100.0%; Pred.No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  218 AACACTTACTTTGAAAA 235
      |||||
Db  54603 AACACTTACTTTGAAAA 54620

RESULT 11
LOCUS      HSJ437116
DEFINITION Human DNA sequence from clone 437116 on chromosome 1p35.1-35.3,
ACCESSION  AL049570
            complete sequence.
NID        94938316
VERSION    AL049570.11 GI:4938316
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 120385)
AUTHORS    Howden,P.
TITLE      Direct Submission
JOURNAL    Submitted (09-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            Requests: clonerequests@sanger.ac.uk
            On Jun 1, 1999 this sequence version replaced gi:4757070.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL
            This sequence is the entire insert of clone 437116. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.

This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
437116 is from the library RPCI3 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR: pCRPC2.

FEATURES
SOURCE
Location/Qualifiers
1..120385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="437116"
/map="p35.1-35.3"
/clone_lib="RPCI3"

```

repeat_region	140. .200	/note="12 repeat: matches 2645. .2710 of consensus"
repeat_region	418. .593	/note="MIR repeat: matches 12. .196 of consensus"
repeat_region	1031. .1090	/note="4 copies 15 mer 90% conserved"
repeat_region	1039. .1328	/note="145 copies 2 mer cc 56% conserved"
repeat_region	1183. .1242	/note="4 copies 15 mer 80% conserved"
repeat_region	4303. .4475	/note="MIR repeat: matches 46. .247 of consensus"
repeat_region	5294. .5442	/note="12 repeat: matches 2557. .2701 of consensus"
repeat_region	5572. .5791	/note="Alu repeat: matches 83. .300 of consensus"
repeat_region	5603. .6094	/note="Alu repeat: matches 3. .289 of consensus"
repeat_region	6173. .6468	/note="12 repeat: matches 2442. .2750 of consensus"
repeat_region	7195. .7240	/note="73 copies 2 mer tg 80% conserved"
repeat_region	7245. .7390	/note="73 copies 2 mer gt 67% conserved"
repeat_region	8220. .8428	/note="MIR repeat: matches 27. .254 of consensus"
repeat_region	10333. .10536	/note="MIR repeat: matches 11. .225 of consensus"
repeat_region	10786. .10960	/note="Alu repeat: matches 153. .309 of consensus"
repeat_region	10961. .11255	/note="Alu repeat: matches 1. .297 of consensus"
repeat_region	11256. .11402	/note="Alu repeat: matches 1. .153 of consensus"
repeat_region	11450. .11749	/note="Alu repeat: matches 3. .302 of consensus"
repeat_region	11804. .13000	/note="L1M3 repeat: matches 4847. .5950 of consensus"
repeat_region	13001. .13322	/note="Alu repeat: matches 1. .312 of consensus"
repeat_region	13323. .13667	/note="L1M3 repeat: matches 4530. .4847 of consensus"
repeat_region	13668. .13968	/note="Alu repeat: matches 1. .297 of consensus"
repeat_region	13969. .14478	/note="L1M3 repeat: matches 3768. .4530 of consensus"
repeat_region	14479. .14799	/note="Alu repeat: matches 1. .306 of consensus"
repeat_region	14800. .14828	/note="L1M3 repeat: matches 3743. .3768 of consensus"
repeat_region	14843. .14996	/note="L1M repeat: matches 3323. .3461 of consensus"
repeat_region	15031. .15329	/note="Alu repeat: matches 4. .301 of consensus"
repeat_region	15342. .15454	/note="MIR repeat: matches 41. .155 of consensus"
repeat_region	15955. .16267	/note="Alu repeat: matches 2. .309 of consensus"
repeat_region	16720. .16885	/note="MIR repeat: matches 23. .196 of consensus"
repeat_region	17086. .17184	/note="MIR repeat: matches 20. .147 of consensus"
repeat_region	17983. .18227	/note="Alu repeat: matches 1. .262 of consensus"
repeat_region	18524. .18571	/note="MIR repeat: matches 91. .141 of consensus"
repeat_region	18583. .18738	/note="L1M3 repeat: matches 280. .435 of consensus"
repeat_region	18904. .19038	/note="L2 repeat: matches 2359. .2488 of consensus"
repeat_region	19185. .19372	/note="Alu repeat: matches 1. .375 of consensus"
repeat_region	19573. .19991	



```

repeat_region /note="LPA5 repeat: matches 5725. .6143 of consensus"
19992. .20527
/note="Tiger3b repeat: matches 375. .892 of consensus"
repeat_region 20528. .20826
/note="AluSg repeat: matches 1. .298 of consensus"
repeat_region 20827. .21089
/note="Tiger3b repeat: matches 892. .1228 of consensus"
repeat_region 22938. .23153
/note="MIR repeat: matches 21. .254 of consensus"
repeat_region 23725. .23835
/note="L2 repeat: matches 2173. .2284 of consensus"
repeat_region 23857. .24103
/note="MIR repeat: matches 23. .262 of consensus"
repeat_region 24307. .24718
/note="L2 repeat: matches 2309. .2750 of consensus"
repeat_region 25147. .25234
/note="44 copies 2 mer tc 65% conserved"
repeat_region 25728. .26034
/note="AluSg repeat: matches 1. .311 of consensus"
repeat_region 26098. .26292
/note="AluJo repeat: matches 126. .302 of consensus"
repeat_region 26293. .26564
/note="AluJb repeat: matches 29. .302 of consensus"
repeat_region 26565. .26604
/note="AluJo repeat: matches 85. .126 of consensus"
repeat_region 26720. .26951
/note="AluJo repeat: matches 1. .227 of consensus"
repeat_region 27111. .27176
/note="MIR repeat: matches 103. .169 of consensus"
repeat_region 27202. .27334
/note="MERSA repeat: matches 27. .187 of consensus"
repeat_region 27534. .27760
/note="MIR repeat: matches 35. .262 of consensus"
repeat_region 30320. .30478
/note="MIR repeat: matches 86. .245 of consensus"
repeat_region 31432. .31551
/note="LIME3A repeat: matches 6039. .6157 of consensus"
repeat_region 31552. .31842
/note="AluSg1 repeat: matches 1. .289 of consensus"
repeat_region 31862. .32397
/note="LIME repeat: matches 5477. .6053 of consensus"
repeat_region 32562. .32856
/note="AluSx repeat: matches 1. .294 of consensus"
repeat_region 34138. .34181
/note="22 copies 2 mer tg 86% conserved"
repeat_region 34303. .34476
/note="MIR repeat: matches 4. .178 of consensus"
repeat_region 35363. .35520
/note="MIR repeat: matches 66. .232 of consensus"
repeat_region 35597. .35675
/note="MIR repeat: matches 109. .182 of consensus"
repeat_region 38012. .38364
/note="L2 repeat: matches 2347. .2746 of consensus"
repeat_region 38856. .38974
/note="MIR repeat: matches 14. .147 of consensus"
repeat_region 39001. .39318
/note="AluJb repeat: matches 1. .301 of consensus"
repeat_region 39415. .39542
/note="MIR repeat: matches 7. .131 of consensus"
repeat_region 39578. .39760
/note="MIR repeat: matches 57. .238 of consensus"
repeat_region 40245. .40575
/note="L2 repeat: matches 2317. .2661 of consensus"
repeat_region 40751. .40784
/note="17 copies 2 mer tg 85% conserved"
repeat_region 44632. .44750
/note="MIR repeat: matches 93. .212 of consensus"
repeat_region 45153. .45222
/note="MIR repeat: matches 79. .150 of consensus"
repeat_region 45433. .45519
/note="MIR repeat: matches 47. .135 of consensus"
repeat_region 45784. .46054
/note="MIR repeat: matches 100. .381 of consensus"

```

```

repeat_region 46232. .46320
/note="L2 repeat: matches 2599. .2694 of consensus"
repeat_region 46520. .46618
/note="MIR repeat: matches 149. .248 of consensus"
repeat_region 46695. .46908

Query Match 7.3% Score 18; DB 10; Length 120385;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 201 TTTTGTTAACAGAA 218
Db 18385 TTTTGTTAACAGAA 18402

RESULT 12
AC003013 LOCUS AC003013 173452 bp DNA 20-OCT-1997
DEFINITION Human PAC clone DJ0205E24 from Xq23, complete sequence.
ACCESSION AC003013
NID 92547257
VERSION AC003013.1 GI:2547257
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (Bases 1 to 173452)
AUTHORS Zhu,H., Tin-Mollam,A and Ozersky,P.
TITLE The sequence of H. sapiens PAC clone DJ0205E24
REFERENCE Unpublished (1997)
AUTHORS 2 (Bases 1 to 173452)
Waterson,R.
DIRECT SUBMISSION Submitted (20-OCT-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by David Bentley's chromosome X mapping group at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK. Further information can be found at <http://www.sanger.ac.uk/HGP/chrX>

SOURCE INFORMATION: This clone was derived from human PAC library RPI-1 prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. For further details, see <http://bacpac.med.buffalo.edu/>. The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>). VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of DJ0205E24;  
actual end is at 173452 of DJ0205E24. The orientation of this clone  
is unknown.

This clone contains STS SWXD2788 (NTD:g1256092).

Location/Qualifiers  
1. 173452

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="DJ0205E24"  
/map\_lib="RRC1-1"  
/map="Xq23"  
repeat\_region complement(402. .692)  
/rpt\_family="ALU"  
repeat\_region complement(831. .1119)  
/rpt\_family="ALU"  
2505. .2787  
repeat\_region complement(3769. .3821)  
/rpt\_family="ALU"  
5716. .6101  
repeat\_region complement(11877. .12822)  
/rpt\_family="L1"  
6137. .6428  
repeat\_region complement(12531. .12822)  
/rpt\_family="L1"  
6433. .6548  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
6555. .6731  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
6743. .6876  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
7042. .7109  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
7196. .7354  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
8058. .8483  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
8648. .9491  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
11549. .11877  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
12822. .12822  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
13655. .13733  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
14704. .14994  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
15002. .15239  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
15296. .15568  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
15573. .15858  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
15862. .16612  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
16665. .16953  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
17371. .17371  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
18844. .19022  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
19047. .19335  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"  
20545. .20545  
repeat\_region complement(13640. .13733)  
/rpt\_family="L1"  
24437. .24727  
repeat\_region complement(13640. .13733)  
/rpt\_family="ALU"

repeat\_region 26757. .27049  
/rpt\_family="ALU"  
repeat\_region complement(27369. .27541)  
/rpt\_family="L1"  
27542. .27833  
repeat\_region complement(27840. .27964)  
/rpt\_family="L1"  
repeat\_region complement(31445. .31468)  
/rpt\_family="L1"  
31491. .31519  
repeat\_region complement(32002. .32293)  
/rpt\_family="L1"  
32698. .32988  
repeat\_region complement(34115. .34167)  
/rpt\_family="ALU"  
34276. .34276  
repeat\_region complement(34322. .34596)  
/rpt\_family="ALU"  
35124. .35273  
repeat\_region complement(35736. .35845)  
/rpt\_family="ALU"  
35736. .35845  
repeat\_region complement(39038. .39082)  
/rpt\_family="ALU"  
39729. .41116  
repeat\_region complement(42581. .42869)  
/rpt\_family="L1"  
42581. .42869  
repeat\_region complement(43574. .43862)  
/rpt\_family="ALU"  
43574. .43862  
repeat\_region complement(45677. .45712)  
/rpt\_family="L1"  
45677. .45712  
repeat\_region complement(47182. .47210)  
/rpt\_family="L1"  
47182. .47210  
repeat\_region complement(48718. .49000)  
/rpt\_family="ALU"  
48718. .49000  
repeat\_region complement(49458. .49757)  
/rpt\_family="ALU"  
49458. .49757  
repeat\_region complement(56722. .57015)  
/rpt\_family="ALU"  
56722. .57015  
repeat\_region complement(58365. .58488)  
/rpt\_family="L1"  
58365. .58488  
repeat\_region complement(58785. .58849)  
/rpt\_family="ALU"  
58785. .58849  
repeat\_region complement(59517. .59699)  
/note="similar to EST AA356244 (NTD:g2008716)"  
59517. .59699  
repeat\_region complement(61376. .61671)  
/rpt\_family="ALU"  
61376. .61671  
repeat\_region complement(63152. .63152)  
/rpt\_family="L1"  
63152. .63152  
repeat\_region complement(64081. .64117)  
/rpt\_family="L1"  
64081. .64117  
repeat\_region complement(64136. .64160)  
/rpt\_family="L1"  
64136. .64160  
repeat\_region complement(66391. .66391)  
/rpt\_family="L1"  
66391. .66391  
repeat\_region complement(67493. .67519)  
/rpt\_family="L1"  
67493. .67519  
repeat\_region complement(68771. .68791)  
/rpt\_family="L1"  
68771. .68791  
repeat\_region complement(72775. .72931)  
/note="similar to EST AA151971 (NTD:g1720827) z030b03.r1"  
72775. .72931  
repeat\_region complement(73014. .73014)  
/note="similar to EST H70965 (NTD:g1042781) yr73e03.r1"  
73014. .73014  
repeat\_region complement(73919. .73919)  
/note="similar to EST AA055592 (NTD:g1547931) z174c12.s1"  
73919. .73919  
repeat\_region complement(72975. .73194)  
/note="similar to EST AA055592 (NTD:g1547931) z174c12.s1"  
72975. .73194

Query Match

7.3%; Score 18; DB 11; Length 173452;



```
repeat_region complement(17095..17249)
repeat_region /rpt_family="L2"
repeat_region /rpt_family="Aluv"
repeat_region complement(17680..17978)
repeat_region /rpt_family="AlusX"
repeat_region complement(18029..18264)
repeat_region /rpt_family="L2"
repeat_region complement(18501..18557)
repeat_region /rpt_family="L2"
repeat_region 18862..19011
repeat_region /rpt_family="Aluub"
repeat_region 19133..19264
repeat_region /rpt_family="MER3"
repeat_region 19402..19468
repeat_region /rpt_family="MER33"
repeat_region complement(19730..19775)
repeat_region /rpt_family="L2"
repeat_region 19790..20090
repeat_region /rpt_family="AlusG"
repeat_region complement(20999..21058)
repeat_region /rpt_family="MIR"
repeat_region complement(21114..21260)
repeat_region /rpt_family="MIR"
repeat_region 21261..21583
repeat_region /rpt_family="MSTA"
repeat_region 21385..21815
repeat_region /rpt_family="AlusX"
repeat_region 21828..21898
repeat_region /rpt_family="MSTA"
repeat_region complement(22740..22807)
repeat_region /rpt_family="(CAAA)n"
repeat_region complement(22811..23111)
repeat_region /rpt_family="AlusX"
repeat_region 23116..23651
repeat_region /rpt_family="MT1F"
repeat_region 23979..24102
repeat_region /rpt_family="MIR"
repeat_region 24355..24895
repeat_region /rpt_family="MER41C"
repeat_region complement(24995..25377)
repeat_region /rpt_family="MER39"
repeat_region complement(25395..25568)
repeat_region /rpt_family="FRAM"
repeat_region complement(26249..26554)
repeat_region /rpt_family="Aluub"
repeat_region 27046..27410
repeat_region /rpt_family="THE1C"
repeat_region 27411..27650
repeat_region /rpt_family="THE1C-internal"
repeat_region complement(27659..27787)
repeat_region /rpt_family="MIR"
repeat_region complement(28143..28323)
repeat_region /rpt_family="MIR"
repeat_region complement(28733..28753)
repeat_region /rpt_family="AT_rich"
STS 29380..29570
/standard_name="CHLC.GCT12610, Chr. -, Homo sapiens"
/db_xref="dbSTS:16208"
repeat_region complement(29471..29547)
repeat_region /rpt_family="MIR"
repeat_region 30115..30206
repeat_region /rpt_family="TTR16C"
repeat_region complement(30604..30903)
```

Query Match 7.3%; Score 18; DB 11; Length 181098;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 107 TACAGATTTCTGATCT 124  
|||||  
DB 85067 TACAGATTTCTGATCT 85050

```
RESULT 14
RATINTLAA 4593 bp DNA ROD 18-NOV-1991
LOCUS Rat alpha-Internexin gene, complete cds.
DEFINITION M73048
ACCESSION M73048
NID 9204963
VERSION M73049.1 GI:204963
KEYWORDS alpha-Internexin.
SOURCE Rattus norvegicus DNA.
ORANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 4593)
AUTHORS Ching/G.Y. and Liem,R.K.H.
TITLE Structure of the gene for the neuronal intermediate filament
protein alpha-Internexin and functional analysis of its promoter
JOURNAL J Biol. Chem. 266, 19459-19468 (1991)
MEDLINE 92011743
FEATURES
source Location/Qualifiers
1..4593
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..4535
/partial
/product="alpha-Internexin"
TATA_signal 1192..1197
exon 1220..2357
/number=1
/product="alpha-Internexin"
join(1293..2357,2358..2482,2483..2810)
CDS
/codon_start=1
/product="alpha-Internexin"
/protein_id="AAA4144.1"
/db_xref="pid:9204964"
/db_xref="gi:204964"
/translation="MSRGSSEHYICSAASYRYKVGDSRLSARLSGPGASGSRSSLS
RSNVASTACSSASSIGLGLAYRRLPASGLDISQAAKNTKIIIRINERQDGLN
DRFAVPIEVHOLETONRALLEALRORHAPSPVGLFQRELHELRAOLESSA
RAQALLERDGLAEVORLARCEESRGREGAERAAQOVDGATLARLEKVE
SLIDELAFVROVDEVEVAILLATIQASOAAEVDAVAKPDLSLRIRQYESLA
AKVLSAEEMYSKSPANLNQOARSPFAIRASSEEHEHYRQLOARTIEIGRGNE
SLEROLLEERHSAVAGVDSIGLESDIRTKSEMAHRLREYDILNVMAADIE
IAAYRKLIAGEEIRFESTSGLSGLPLNPSTLPRLSLSTTSKVSAGSLKKE
EESEEEGASKEVTKTSKVGSEFETLVSTKTEKSTIEITTSQKM"
exon 2358..2482
/number=2
/product="alpha-Internexin"
2483..4556
/number=3
/exon
BASE COUNT 1159 a 1273 c 1187 g 974 t
ORIGIN
Query Match 7.3%; Score 18; DB 12; Length 4593;  
Best Local Similarity 100.0%; Pred. No. 9.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 198 GTCCTTTTGTAAACCA 215  
|||||  
DB 491 GTCCTTTTGTAAACCA 508

RESULT 15  
AC005050/c 186882 bp DNA HTG 12-JUN-1998  
LOCUS Homo sapiens clone R6030L05, WORKING DRAFT SEQUENCE, 3 unordered  
DEFINITION pieces.  
ACCESSION AC005050  
NID 93212941  
VERSION AC005050.1 GI:3212941  
KEYWORDS HTG; HTGS\_PHASE1.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 188682)  
 AUTHORS Waterston, R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 188682)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

COMMENT  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1496: contig of 1496 bp in length  
 \* 1497 1513: gap of unknown length  
 \* 1514 72532: contig of 71019 bp in length  
 \* 72533 72549: gap of unknown length  
 \* 72550 188682: contig of 116133 bp in length.

FEATURES  
 source  
 1. 188682  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RG030L05"

BASE COUNT 57647 a 38033 c 37935 g 55033 t 34 others  
 ORIGIN

Query Match 7.3% Score 18: DB 34; Length 188682;  
 Best Local Similarity 100.0%; Fred. No. 11;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 175 CAAGGAGCCCAACTCTG 192  
 ||||||||||||||||  
 Db 39753 CAAGGAGCCCAACTCTG 39736

Search completed: August 6, 1999, 11:42:09  
 Job time: 6615 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:40:51; Search time 650.92 Seconds

(without alignments)  
87,251 Million cell updates/sec

Title: US-09-049-696-13

Perfect score: 227  
Sequence: 1 GTGGCGGCTCTGGAGAGT.....AATGCTCCACTGATCT 227

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N.Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	227	100.0	878	T45884	Human colon specif
2	227	100.0	878	V16672	Polynucleotide seq
3	17	7.5	2129	O63972	Megakaryocyte pote
4	17	7.5	80073	T58840.5	Continuation (6 of
5	17	7.5	2138	T91079	Human CARK antigen
6	17	7.5	2454	T97392	Aspergillus oryzae
7	16	7.0	261	O61061	Human brain Expres
8	16	7.0	3600	T27535	Invasin protein ge
9	16	7.0	1851	T78543	Human mitogen-acti
10	16	7.0	1853	T90352	Human MAP kinase h
11	16	7.0	476	T83728	DNA encoding a Sta
12	16	7.0	1298	V26081	Human stress-activ
13	16	7.0	14654	V52239	Streptococcus pneu
14	16	7.0	9223	V52192	Streptococcus pneu
15	16	7.0	110000	V21209.14	Continuation (15 o
16	16	7.0	1851	V33561	Human MAP kinase c
17	16	7.0	6060	V74845	Staphylococcus aur
18	16	7.0	9956	V74663	Staphylococcus aur
19	16	7.0	3516	V74546	Staphylococcus aur
20	16	7.0	1838	X01086	Human CSBP beta po
21	16	7.0	5194	X25885	C. albicans alpha-I
22	15	6.6	1606	N90781	Sequence of a cDNA
23	15	6.6	1584	N90783	Sequence of a cDNA
24	15	6.6	3086	N91815	Egr-1 contig. OC3.1
25	15	6.6	1730	N80433	Bovine Interleukin
26	15	6.6	1606	O06415	Sequence encoding
27	15	6.6	2184	O06301	Tosoplasma gondii
28	15	6.6	1909	O11919	Murine Interleukin
29	15	6.6	1606	O69157	Soluble human inte
30	15	6.6	735	O73229	Egr-1 transcriptio
31	15	6.6	843	O73457	Mouse Egr-1 clone
32	15	6.6	3086	O73463	DNA encoding cytop
33	15	6.6	1942	O84888	Megakaryocyte kina
34	15	6.6	2000	T00616	Human gene signatu
35	15	6.6	276	T19580	Human gene signatu
36	15	6.6	267	T25793	Human gene signatu
37	15	6.6	2255	T44510	E. coli O157:H7 FI
38	15	6.6	60	T51375	3' primer for huma
39	15	6.6	960	T39047	CDNA encoding cell
40	15	6.6	894	T39061	Chimeric endogluc
41	15	6.6	927	T39062	Chimeric endogluc
42	15	6.6	1308	T64453	PVR protease gene.
43	15	6.6	1909	T59277	DNA encoding T. go

44 15 6.6 2075 1 T75444  
c 45 15 6.6 1441 1 X37424

CDNA encoding oste  
Human secreted pro

## ALIGNMENTS

```
RESULT 1
ID T45884
AC T45884 standard; cDNA; 878 BP.
DT 13-MAR-1997 (first entry)
DE Human colon specific gene CSG5 cDNA partial clone.
KW Human colon specific gene; CSG5; colon cancer; metastasis; diagnosis;
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 2..691
FT 1 /*tag= a
FT FT
FT WO9639419-A1.
PD 12-DEC-1996.
PE 06-JUN-1995; U07289.
PR 06-JUN-1995; WO-U07289.
PA (HOMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Yu G;
PI WPI; 97-043054/04.
DR P-PSDB; W06548.
PT Human colon specific genes and their expression products - detection
PT of which, in non-colon tissue samples, can be used as indication of
PT colon cancer metastasis
PS Claim 1; Fig 5; 60pp; English.
CC 13 CDNA clones (T45880-92), most of them partial clones, correspond
CC to human colon specific genes, designated CSG1, CSG2, etc., that
CC are primarily expressed in tissues derived from the colon. CSG7
CC and CSG10 show reduced expression in colon cancer cells as compared
CC to that in normal cells; the remaining genes are overexpressed in
CC colon cancer. The partial cDNA sequences can be used to isolate
CC full-length clones and genomic clones including the complete gene.
CC CSG nucleic acids can be used to produce CSG polypeptides (see also
CC W06545-53) in transformed host cells, as probes to detect disorders
CC of the colon, partic. colon cancer and colon cancer metastasis, and
CC in gene therapy.
SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;
```

Query Match 100.0%; Score 227; DB 1; Length 878;  
Best Local Similarity 100.0%; Pred. No. 5, 1e-110;  
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTGGCGGCTCTGGAGAGATTACGACGACGAGAGAGATATACCCAGCAGAGTGA 60
DB 56 GTGGCGGCTCTGGAGAGATTACGACGACGAGAGAGATATACCCAGCAGAGTGA 115
QY 61 GCACGTGATACCTCGCTCGATTGAGATGTAATGAAATCAATGAAATCCACCAAGACT 120
DB 116 GCACGTGATACCTCGCTCGATTGAGATGTAATGAAATCAATGAAATCCACCAAGACT 175
QY 121 GAATTAATTAAGATGATGTTCAACACAGCAAGAGTGTTCAGACAGAAATCCCGGGA 180
DB 176 GAATTAATTAAGATGATGTTCAACACAGCAAGAGTGTTCAGACAGAAATCCCGGGA 235
QY 181 GGCTCATTTGGCTTCGATGTCCCAATGCTCCCATACCTGATCT 227
DB 236 GGCTCATTTGGCTTCGATGTCCCAATGCTCCCATACCTGATCT 282
```

RESULT 2  
ID V16672  
AC V16672 standard; cDNA; 878 BP.  
DT 22-JUN-1998 (first entry)  
DE Polynucleotide sequence of a colon-specific gene.  
KW Colon-specific gene; probe; detection; expression; human;

KW diagnostic assay; colon cancer; antibody; screening; ss.  
OS Homo sapiens. Location/Qualifiers  
FH Key 2..685  
FT CDS  
FT /tag= a  
FT /note= "no stop codon given"  
PI US5733748-A.  
PI 31-MAR-1998.  
PI 06-JUN-1995; 469667.  
PI 06-JUN-1995; US-469667.  
PI (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen C, Yu G.  
PI WPI: 98-229823/20.  
PI P-PSDB: W46879.  
PI Colon-specific nucleic acids - useful as probes for detecting colon  
PI cancer micrometastases  
PS Claim 15; Fig 5A-B; 51pp; English.  
CC V16668-81 represent polynucleotide sequences of partial or full-length  
CC cDNA clones of colon-specific genes. The polynucleotides can be used  
CC as probes to detect expression of the corresponding human genes, e.g. in  
CC diagnostic assays for detecting micrometastases of colon cancer.  
CC Recombinant cells containing the polynucleotides can be used to  
CC produce the polypeptides, in order that antibodies can be raised and  
CC used in further screening or diagnostics.  
CC Sequence 878 BP: 257 A; 179 C; 188 G; 241 T;  
SQ

Query Match 100.0%; Score 227; DB 1; Length 878;  
Best Local Similarity 100.0%; Pred. No. 5,1e-110;  
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCGGGCTCTGGGAGGATTACCGACGAGAGTGATACCCGACGAGAGTGA 60  
DB 56 GTGCGGGCTCTGGGAGGATTACCGACGAGAGTGATACCCGACGAGAGTGA 115  
QY 61 GCACGTACATACCTGGGTGATGAGATGTAATGCAATGCAATCCACCAAGACT 120  
DB 116 GCACGTACATACCTGGGTGATGAGATGTAATGCAATGCAATCCACCAAGACT 175  
QY 121 GAAATTAATAGAGATGATGTCACACAGCAAGTGTGTTCCAGAGAACTCTCGGGA 180  
DB 176 GAAATTAATAGAGATGATGTCACACAGCAAGTGTGTTCCAGAGAACTCTCGGGA 235  
QY 181 GGCTATTGCTGGCTTGTGATGTCCCAATGCTCCCATACCTGATCT 227  
DB 236 GGCTATTGCTGGCTTGTGATGTCCCAATGCTCCCATACCTGATCT 282

RESULT 3  
063972  
ID 063972 standard; cDNA; 2129 BP.  
AC 063972;  
DT 08-DEC-1994 (first entry)  
DE Megakaryocyte potentiator (PKP027).  
KW Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;  
KW platelet; ss.  
OS Homo sapiens.  
FH Key  
FT cds Location/Qualifiers  
FT /tag= a  
FT /tag= b  
FT /product= Meg-Pot  
FT /note= "claim 1; 584 amino acid product"  
FT /tag= c  
FT /note= "claim 6; 248 amino acid product"  
PI WO9410312-A.  
PI 11-MAY-1994.  
PI 25-OCT-1993; J01540.  
PI 23-OCT-1992; JP-286153.  
PI 11-NOV-1992; JP-301387.  
PI 09-DEC-1992; JP-325546.

PA (CHUS) CHUGAI SEIYAKU KK.  
PI Hattori K, Kojima T, Oh-eda M, Yamaguchi N;  
DR WPI: 94-167467/20.  
DR P-PSDB: R53992.  
PI New mega:karyocyte potentiator - for potential treatment of  
PI thrombocytopenia  
PS Disclosure; page 52-57; 74pp; Japanese.  
CC DNA encoding the 584 or 248 amino acid prod. has potential  
CC use in treatment of thrombocytopenia and low platelet function.  
SQ Sequence 2129 BP; 389 A; 715 C; 658 G; 367 T;

Query Match 7.5%; Score 17; DB 1; Length 2129;  
Best Local Similarity 100.0%; Pred. No. 6.8;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCGGGCTCTGGGAGG 17  
DB 640 GTGCGGGCTCTGGGAGG 656

RESULT 4  
T58840\_5  
Continuation (6 of 6) of T58840 from base 500001 (Mycoplasma genitalium genome.)  
WP Sequence split into 6 fragments LOCUS T58840 Accession T58840  
WP Fragment Name Begin End  
WP T58840\_0 1 110000  
WP T58840\_1 100001 210000  
WP T58840\_2 200001 310000  
WP T58840\_3 300001 410000  
WP T58840\_4 400001 510000  
WP T58840\_5 500001 580073

Query Match 7.5%; Score 17; DB 1; Length 80073;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 TCAACACAGCAAGTGT 157  
DB 39671 TCAACACAGCAAGTGT 39687

RESULT 5  
T91079  
ID T91079 standard; cDNA; 2138 BP.  
AC T91079;  
DT 11-MAR-1998 (first entry)  
DE Human CAK1 antigen (mesothelin) cDNA.  
KW CAK1 antigen; mesothelin; tumour specific antigen; mesothelioma;  
KW ovarian cancer; squamous cell cancer; gene therapy; diagnosis; ss.  
OS Homo sapiens.  
FH Key  
FT CDS Location/Qualifiers  
FT /tag= a  
FT /tag= b  
FT /note= "putative signal sequence for membrane  
FT /note= "insertion"  
FT /tag= c  
FT /note= "variant polyadenylation signal"  
PI WO9725068-A2.  
PI 17-JUL-1997.  
PI 03-JAN-1997; U00224.  
PI 05-JAN-1996; US-010166.  
PI (USSH) US DEPT HEALTH & HUMAN SERVICES.  
PI Chang K, Pastan I;  
DR WPI: 97-372620/34.  
DR P-PSDB: W26674.  
PI Agents for targeting mesothelin, a tumour cell antigen - used for  
PI the detection or inhibition of growth of e.g. mesotheliomas, ovarian  
PI cancers and squamous cell carcinomas



PS Claim 16; Page 58-60; 72pp; English.  
CC This cDNA clone codes for CAK1 antigen (see W26674) which is  
CC found on mesothelium, mesotheliomas, ovarian cancers and some  
CC squamous cell carcinomas. The antigen has been designated  
CC mesothelin. The clone was isolated from a HeLa S3 cDNA library by  
CC screening with monoclonal antibody K1, previously shown to react  
CC with an antigen on human ovarian tumour OVCAR-3 cells. A 40 kDa  
CC form (K1) of mesothelin appears to be derived from the 69 kDa  
CC precursor by several processing steps. A claimed method for  
CC specifically delivering an effector molecule to a tumour cell  
CC bearing an antigen comprising at least 10 contiguous amino acids  
CC of mesothelin involves: (a) providing a chimeric molecule comprising  
CC the effector molecule attached to a targeting molecule that  
CC specifically binds to mesothelin; and (b) contacting the tumour with  
CC the chimeric molecule such that the chimeric molecule specifically  
CC binds to a tumour cell. Also claimed is a method for inhibiting  
CC mesothelin expression or activity by contacting mesothelin bearing  
CC cells with inhibitory nucleic acids for the mesothelin gene. The  
CC methods can be used to detect tumour cells and to inhibit the  
CC growth of cells bearing mesothelin. Mesothelin-derived antigens  
CC may be used in vaccines for the inhibition or prevention of  
CC mesotheliomas or ovarian tumours. DNA encoding the mesothelin  
CC antigen can be transfected into a mammal containing tumour cells to  
CC screen for drugs useful in cancer treatment. Antisense  
CC oligonucleotides can be used to inhibit mesothelin expression.  
CC (All claimed).  
SQ Sequence 2138 BP; 385 A; 722 C; 661 G; 370 T;

Query Match 7.5%; Score 17; DB 1; Length 2138;  
Best Local Similarity 100.0%; Pred. No. 6.8;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCGGCTCTGGAGG 17  
|||||  
DB 637 GTGCGGCTCTGGAGG 653

RESULT 6  
T97392/c  
ID T97392 standard; DNA: 2454 BP.  
AC T97392:  
DT 11-MAY-1998 (first entry)  
DE Aspergillus oryzae protease pepc gene.  
KW pepc gene; protease; pepc; Area regulator; gene replacement;  
KW antisense; gene inactivation; protease-free host cell;  
KW recombinant protein; ss.  
OS Aspergillus oryzae strain IFO 4178.  
FH Key Location/Qualifiers  
FT CDS  
FT /tag= a  
FT /note= "Contains 3 introns"  
FT /tag= b  
FT /number= 1  
FT /tag= c  
FT /number= 1  
FT /tag= d  
FT /number= 2  
FT /tag= e  
FT /number= 2  
FT /tag= f  
FT /number= 3  
FT /tag= g  
FT /number= 3  
FT /tag= h  
FT /number= 4

PN W09722705-A1.  
PD 26-JUN-1997.  
PF 16-DEC-1996; DR0528.  
PR 15-DEC-1995; DR-001428.  
PA (NOVO) NOVO-NORDISK AS.  
PI Christensen T, Lehmebeck J;  
DR WPI: 97-341696/31.  
P-PSDB: W31628.  
PT New fungus lacking functional Area and extracellular protease  
PT gene(s) - for high yield expression of industrial or therapeutic  
PT proteins, also new protease(s) from Aspergillus oryzae  
PS Example 1; Page 31-33; 80pp; English.  
CC This DNA sequence includes a coding region for the aspartic-type  
CC protease Pepc (see W31628) of Aspergillus oryzae. The pepc gene  
CC was cloned by cross-hybridisation with the Aspergillus niger pepc  
CC gene. The invention relates to a new fungus in which: (a) the area  
CC gene (see T97394) has been modified by recombinant DNA methods so  
CC that it cannot express a functional Area activator; and (2) the  
CC extracellular protease pepc (see T97393) and/or pepc genes are  
CC inactivated so that functional enzyme is not produced. The  
CC fungi have complete or partial deletions of the appropriate genes,  
CC or expression sequences that regulate the genes. Alternatively,  
CC the genes are inactivated by antisense mechanisms or by gene  
CC replacement. The new fungi are used for the recombinant production  
CC of industrial or therapeutic peptides and proteins, particularly  
CC enzymes (e.g. proteases, lipase, cellulase or chymosin),  
CC insulin, growth hormone, glucagon, somatostatin, interferon,  
CC platelet-derived growth factor, factor VII, factor VIII, urokinase,  
CC tissue plasminogen activator, erythropoietin or thrombopoietin.  
CC The new fungi do not produce protease and so provide higher yields  
CC of recombinant protein because of reduced proteolytic degradation of  
CC the isolated pepc and pepc genes can also be used for production of  
CC recombinant pep proteases.  
SQ Sequence 2454 BP; 574 A; 651 C; 551 G; 678 T;

Query Match 7.5%; Score 17; DB 1; Length 2454;  
Best Local Similarity 100.0%; Pred. No. 6.8;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 GATGATGAATCAAT 103  
|||||  
DB 514 GATGATGAATCAAT 498

RESULT 7  
O61061/c  
ID O61061 standard; DNA: 261 BP.  
AC O61061:  
DT 16-MAR-1994 (first entry)  
DE Human brain Expressed Sequence Tag EST01682.  
KW Gene transcription product; genetic markers; tagging; in vivo;  
KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
OS Homo sapiens.  
PN W09316178-A.  
PD 19-AUG-1993.  
PF 12-FEB-1993; U01294.  
PR 12-FEB-1992; US-837195.  
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.  
PI Adams MD, Moreno RF, Venter CJ;  
DR WPI: 93-272882/34.  
PT Enriched oligonucleotides and corresp. sequences - used as  
PT markers for human genes transcribed in-vivo, facilitate tagging  
PT of most human genes  
PS Example 4; Page 433; 500pp; English.  
CC The Expressed Sequence Tag was isolated from a human brain cDNA  
CC library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes  
CC on chromosomes, for individual or forensic identification, for mapping  
CC locations of disease-associated genes, for identification of tissue  
CC type, and for prepn. of antisense sequences, probes and constructs.  
CC EST01682 has a "poor" coding probability as evaluated using the

CC coding-region prediction program CRM. See also Q59041-Q61440.  
 .50 Sequence 261 BP; 93 A; 59 C; 58 G; 50 T.

Query Match 7.0%; Score 16; DB 1; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 22;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 TCATTGTGGCTTCG 199  
 |||||||  
 Db 243 TCATTGTGGCTTCG 228

RESULT 8

T27535 127535 standard; cDNA; 3600 BP.

AC T27535; DT 24-AUG-1996 (first entry)

DE Invasin protein gene.

KW Invasin; drug delivery; gastrointestinal membrane; transcytosis;

KM bioavailability; fusion protein; enterocyte; Peyer's patch M-cell;

OS ss. Versinia enterocolitica strain 8081c.

FH Key Location/Qualifiers

FT cds 413..2920

FT /\*tag= a

FT /product= Invasin protein

PN W09613250-A1.

PD 09-MAY-1996.

PF 20-OCT-1995; U13749.

PR 27-OCT-1994; US-331393.

PA (AMGE-) AMGEN INC.

PI Habberfield AD, Jensen-Piippo K;

DR WPI: 96-251447/25.

P-PSDB; R96206.

PT Therapeutic delivery system utilizing bacterial invasin protein - is

PT not readily degraded in the gut, enhances systemic bio-availability

PT of therapeutic agents

PS Example 1: Fig 1: 110pp; English.

CC The sequence encodes a bacterial invasin protein, which may be

CC complexed with a therapeutic agent to transport the agent across the

CC gastrointestinal membrane barrier by transcytosis to increase

CC bioavailability 5- to 100-fold. The agent and bacterial protein may

CC be linked via a degradable peptide sequence as a fusion protein.

CC The invasin protein gene may be fused with a maltose binding protein

CC gene (T27537) for expression of a fusion protein (R96209) which may

CC be purified easily by amylose affinity chromatography. The delivery

CC system allows improved transport across enterocytes and Peyer's patch

CC M-cells. The system is not prone to degradation in the gut or early

CC release of biologically active material, and eliminates the need

CC for parenteral administration.

CC Sequence 3600 BP; 965 A; 777 C; 842 G; 1016 T;

SO

Query Match 7.0%; Score 16; DB 1; Length 3600;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 92 ATGAATACATGGA 107  
 |||||||  
 Db 1260 ATGAATACATGGA 1275

RESULT 9

T78543/c 178543 standard; cDNA; 1851 BP.

AC T78543; DT 29-OCT-1997 (first entry)

DE Human mitogen-activated protein kinase homologue SMAP encoding cDNA.

KW MAP kinase; gastritis; ulcer; viral infection; bacterial infection;

KM neoplasm; stomach; human; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 59..1154  
 FT /\*tag= a  
 FT /product= SMAP  
 FT /transl\_except= (pos:350,aa:gly)  
 FT /note= "Only one base codes for gly"

PM W09702347-A1.

PD 23-JAN-1997.

PF 28-JUN-1996; U11170.

PR 30-JUN-1995; US-000722.

PA (INCY-) INCYTE PHARM INC.

PI Au-Young J, Guegler KJ, Hawkins PR, Wilde CG;

DR WPI: 97-108960/10.

P-PSDB; W23783.

PT DNA encoding human mitogen-activated protein kinase - used to

PT develop prods. for the diagnosis and treatment of, e.g. gastritis,

PT ulcers, viral and bacterial infections or neoplasms

PS Claim 2: Page 26-27; 41pp; English.

CC The present sequence encodes the novel human mitogen-activated protein

CC (MAP) kinase homologue designated SMAP. The cDNA encoding SMAP was

CC isolated from a human stomach cDNA library. The SMAP protein is

CC specifically involved with protective cell signalling processes and

CC the products can be used in the study, diagnosis and treatment of

CC conditions which affect the stomach such as gastritis, ulcers, viral

CC and bacterial infections, or neoplasms associated with abnormal signal

CC transduction.

CC Sequence 1851 BP; 434 A; 501 C; 523 G; 393 T;

SO

Query Match 7.0%; Score 16; DB 1; Length 1851;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 45 ACCCCAGCAGATGGA 60  
 |||||||  
 Db 493 ACCCCAGCAGATGGA 478

RESULT 10

T90352/c 190352 standard; cDNA; 1853 BP.

AC T90352; DT 20-JAN-1998 (first entry)

DE Human MAP kinase homologue SMAP cDNA.

KW SMAP; MAP kinase homologue; mitogen-activated protein; human;

KM signal transduction; inflammation; therapy; gastritis; ulcer; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 59..1156

FT /\*tag= a

PN U55663313-A.

PD 02-SEP-1997.

PF 28-JUN-1996; 674612.

PR 28-JUN-1996; US-674612.

PA (INCY-) INCYTE PHARM INC.

PI Au-Young J, Guegler KJ, Hawkins PR, Wilde CG;

DR WPI: 97-447978/41.

P-PSDB; W26578.

PT Nucleic acid encoding mitogen-activated protein kinase homologue -

PT for production of recombinant protein, for treatment of

PT inflammation

PS Claim 2: Column 19-22; 19pp; English.

CC This cDNA sequence codes for SMAP (see W26578) a human MAP kinase

CC homologue that is involved in protective cell signalling processes.

CC It was identified and cloned using Incyte Clone No. 214915 from a

CC stomach cDNA library by PCR amplification (see T90353-54). Also

CC claimed are: an expression vector containing the polynucleotide

CC and a host cell transformed with the vector. The isolated

CC polynucleotide may be used for the production of recombinant SMAP,

CC which may be useful for drug screening, drug design, research,

CC antibody production, etc. The polynucleotide and SMAP, or its

CC variants, are especially used in the diagnosis and treatment of

CC activated or inflamed cells and/or tissues associated with the

CC expression of SMAP, especially for the diagnosis and treatment of

CC conditions that affect the stomach such as gastritis, ulcers, viral  
 CC and bacterial infections and neoplasms.  
 SQ Sequence 1853 BP; 434 A; 523 G; 393 T;

Query Match 7.0%; Score 16; DB 1; Length 1853;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 45 ACCCCAGCAGAGTGGA 60  
 ||||||||||||||||  
 DB 495 ACCCCAGCAGAGTGGA 480

RESULT 11  
 T83728  
 ID T83728 standard; DNA; 476 BP.  
 AC T83728:  
 DT 16-JUL-1998 (first entry)  
 DE DNA encoding a Staphylococcus aureus protein of unknown function.  
 CC Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
 KW Staphylococcal gene; regulatory element; bacterial gene expression;  
 KM vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
 KW toxic shock syndrome; ss.  
 OS Staphylococcus aureus  
 FH Location/Qualifiers  
 FT Key  
 FT CDS 262..378  
 FT /tag= a  
 PN WO9730070-A1.  
 PD 21-AUG-1997.  
 PR 19-FEB-1997; U02318.  
 PR 20-FEB-1996; US-011888.  
 PA (SMK) SMITHKLINE BEECHAM CORP.  
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,  
 PI Pratt JM, Reichard RW, Rosenberg W, Ward JM;  
 DR WPI: 97-424969/39.  
 DR P-PSDB: W27259.  
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used  
 PT to isolate antimicrobial compounds, and in vaccines against S.  
 PT aureus infection  
 PS Claim 9: Page 642; 989pp; English.  
 CC The present sequence encodes a Staphylococcus aureus protein of  
 CC unknown function. The present sequence was isolated from a  
 CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA  
 CC sequence can be used in the construction of ribozymes and antisense  
 CC sequences to control the expression of Staphylococcal genes. The DNA  
 CC sequence is also useful as a source of regulatory elements for the  
 CC control of bacterial gene expression. The encoded protein may be used  
 CC to produce vaccines to enable a host to produce specific antibodies  
 CC with antibacterial action. These vaccines and antibodies would protect  
 CC a host against invasion by S. aureus, and conditions relating to  
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled  
 CC skin syndrome, and toxic shock syndrome.  
 SQ Sequence 476 BP; 160 A; 76 C; 77 G; 116 T;

Query Match 7.0%; Score 16; DB 1; Length 476;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 ATAAAGATGATGTTCA 143  
 ||||||||||||||||  
 DB 269 ATAAAGATGATGTTCA 284

RESULT 12  
 V26081/C  
 ID V26081 standard; CDNA; 1298 BP.  
 AC V26081:  
 DT 29-SEP-1998 (first entry)  
 DE Human stress-activated protein kinase 4 (SAPK4) cDNA.  
 KM Stress-activated protein kinase 4; SAPK4; pituitary gland; asthma;  
 KW stress activated kinase kinase 3; SKK3; rheumatoid arthritis;

KW psoriasis; inflammatory disease; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 23..1120  
 FT /tag= a  
 FT /product= "Human stress-activated protein kinase 4"

PN WO9815618-A1.  
 PD 16-APR-1998.  
 PR 09-OCT-1997; G02779.  
 PR 15-MAY-1997; GB-009781.  
 PR 09-OCT-1996; GB-021096.  
 PA (MEDT-) MEDICAL RES COUNCIL.  
 PI Cohen P, Goedert M;  
 DR WPI: 98-240806/21.  
 DR P-PSDB: W55967.  
 PT New stress-activated protein kinase 4 - useful in drug screening,  
 PT for, e.g. anti-inflammatory, anti-cancer and immuno-suppressing  
 PT agents  
 PS Claim 6: Fig 1; 119pp; English.  
 CC The present claimed sequence represents the human stress-activated  
 CC protein kinase 4 (SAPK4) cDNA isolated from a human pituitary gland  
 CC cDNA library. The invention claims that SAPK4 protein can be useful  
 CC in a screening assay for identifying agents that inhibit its activity  
 CC and/or agents that block its activation through stress activated kinase  
 CC kinase 3 (SKK3). Therefore, the agents identified in the assays may be  
 CC potentially useful for treating inflammatory diseases, e.g. rheumatoid  
 CC arthritis, asthma and psoriasis.  
 SQ Sequence 1298 BP; 302 A; 374 C; 374 G; 248 T;

Query Match 7.0%; Score 16; DB 1; Length 1298;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 45 ACCCCAGCAGAGTGGA 60  
 ||||||||||||||||  
 DB 459 ACCCCAGCAGAGTGGA 444

RESULT 13  
 V52239  
 ID V52239 standard; DNA; 14654 BP.  
 AC V52239:  
 DT 23-OCT-1998 (first entry)  
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:106.  
 CC Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 OS Streptococcus pneumoniae.  
 PN MO9818931-A2.  
 PD 07-MAY-1998.  
 PR 30-OCT-1997; U19588.  
 PR 31-OCT-1996; US-029960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
 PI Kunsch CA, Rosen CA;  
 DR WPI: 98-272225/24.  
 PT Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 PS Claim 1: Page 781-790; 1409pp; English.  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridise to the target sequence and isolating the nucleic acid

CC molecules from the members: or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
SQ Sequence 14654 BP: 4426 A; 3125 C; 2804 G; 4299 T;

Query Match 7.0%; Score 16; DB 1; Length 14654;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AATACATGGAATCC 110  
DB 4689 AATACATGGAATCC 4704

RESULT 14  
V52192  
ID V52192 standard; DNA: 9223 BP.

AC V52192;  
DT 23-Oct-1998 (first entry)  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:59.  
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KM computer readable medium; vaccine; pharmaceutical composition; ds.  
OS Streptococcus pneumoniae.  
PN W09818931.A2.

PD 07-MAY-1998;  
PF 30-OCT-1997; U19588.  
PR 31-OCT-1996; US-029960.  
PA (HUMAN-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
PI Kunsch CA, Rosen CA;  
PI WPI: 98-272225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae

PS Claim 1; Page 514-520; 1409pp; English.

CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridize to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
SQ Sequence 9223 BP: 3229 A; 1375 C; 1885 G; 2734 T;

Query Match 7.0%; Score 16; DB 1; Length 9223;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 GAATACATGGAATC 109  
DB 3038 GAATACATGGAATC 3053

RESULT 15  
V21209\_14  
Continuation (15 of 17) of V21209 from base 1400001 (Methanococcus jannaschii citrula  
WP Sequence Split into 17 Fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 7.0%; Score 16; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GAATTAATAGGATG 136  
DB 31441 GAATTAATAGGATG 31456

Search completed: August 5, 1999, 17:41:00  
Job time: 6295 sec



Matches 227: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCGGGCTCTGGAGAGATTAAACGACGAGAGAGTATACCCAGCAGACTGGA 60  
|||||  
DB 56 GTGCGGGCTCTGGAGAGATTAAACGACGAGAGAGTATACCCAGCAGACTGGA 115  
|||||

QY 61 GCACCTACATACCTGCTGCTGATGAGATGAAATACATGATCCACCAAGCT 120  
|||||  
DB 116 GCACCTACATACCTGCTGCTGATGAGATGAAATACATGATCCACCAAGCT 175  
|||||

QY 121 GAAATTAATAGATGATGTCACACAGCAAGTGTGTTTCAGCAACATCCTCGGGA 180  
|||||  
DB 176 GAAATTAATAGATGATGTCACACAGCAAGTGTGTTTCAGCAACATCCTCGGGA 235  
|||||

QY 181 GGCTCATTTGTGCTTCTGATGTGCCAAATGCTCCATACCTGATCT 227  
|||||  
DB 236 GGCTCATTTGTGCTTCTGATGTGCCAAATGCTCCATACCTGATCT 282  
|||||

RESULT 2  
AF039400 3311 bp mRNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLIC1) mRNA.  
DEFINITION complete cds.  
ACCESSION AF039400  
NID 9409457  
VERSION AF039400.1 GI:4009457  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLIC1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLIC1"  
352. .3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:9409458"  
/db\_xref="GI:4009458"  
/translation="MGPKSVFLLIHLLEGALSNSLIQNNNGYEGIVADPNP  
EDFLIQIKDMVTQASLYLEATGKRFYKNAVLIPEWTIKADYVRKLETKNA  
DVALESTPQNDPEPTQMGNGEGERIHLTPDFAGKRLAEPGAKFVHEMAH  
LRMGVDEYNDEKFLYLSNGRIQAVRCSAGITGNVVKCGGSCYTRKCTFNVTGL  
YEKCEFLVSRQTEKASIMQAQVDSIVERCTEQNNKRNKPNKONKCMNRSTWEYI  
RDSDFKKTPTMTQPNPTFSLIQGRIQVCLVLDKSGSMATGNRLNRLNOAGLEL  
LQVTELSWGVMTFDSAAHVSQSLIQINSGSDPDTAKRLPAASGCTSCISBSA  
FTVIRKRYPTDGSFTVILTDGEDNTISGCRNFKVQSGAIIHTVALGSAQCELEJSK  
MTGGLQTVASQVQNNGLIDAFGLSSGNGAVSQRISQLSKGLTTONSQMMNGTVIV  
DSTVGKDTLELTWTQPPQILMDPDSGQGGGVVDKNTKMAVLQIPGIKAGVTWY  
SLQSSQTLTLVTSRASNAITLPITVTSKTKNTSKFSPPLVYVANIROGASPIILA

STALIESVNGKVTLELLDNGAGADATKDDGYRSRYFTTYDINGRSYKVRALGCVN  
AARRRVIPQOSGALYIPGWIENDEILQNNPREPEINKDDVOHKVCFRISGGSFVS  
DVNPAPIPDLEPPOITDLKAEIHGSLINLWTAPDDYDHGAHKYIIRISITLD  
LRDKFNESLQVNTTALIPKRNASEVFLPEPENTPENGDLFALQAVRVKDLKSTI  
SNIRVSLFLPQIPPTPSPDERSAPCPINHINSTITPGIHLIKMKWIGELQSLA

Query Match  
Best local similarity 100.0%; Score 227; DB 11; Length 3311;  
Matches 227: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 1028 a 692 c 742 g 849 t

ORIGIN

QY 1 GTGCGGGCTCTGGAGAGATTAAACGACGAGAGAGTATACCCAGCAGACTGGA 60  
|||||  
DB 2374 GTGCGGGCTCTGGAGAGATTAAACGACGAGAGAGTATACCCAGCAGACTGGA 2433  
|||||

QY 61 GCACCTACATACCTGCTGCTGATGAGATGAAATACATGATCCACCAAGCT 120  
|||||  
DB 2434 GCACCTACATACCTGCTGCTGATGAGATGAAATACATGATCCACCAAGCT 2493  
|||||

QY 121 GAAATTAATAGATGATGTCACACAGCAAGTGTGTTTCAGCAACATCCTCGGGA 180  
|||||  
DB 2494 GAAATTAATAGATGATGTCACACAGCAAGTGTGTTTCAGCAACATCCTCGGGA 2553  
|||||

QY 181 GGCTCATTTGTGCTTCTGATGTGCCAAATGCTCCATACCTGATCT 227  
|||||  
DB 2554 GGCTCATTTGTGCTTCTGATGTGCCAAATGCTCCATACCTGATCT 2600  
|||||

RESULT 3  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CaCC1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 2826)  
AUTHORS Agnel,M., Verma,T. and Culouscou,J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel,M. and Culouscou,J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des Carrières, Neuilly-Malmaison 92500, France  
FEATURES  
source  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="CaCC1"  
5. .2749  
/gene="CaCC1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="GI:4585469"  
/db\_xref="PID:94585469"  
/translation="MGPKSVFLLIHLLEGALSNSLIQNNNGYEGIVADPNP  
EDFLIQIKDMVTQASLYLEATGKRFYKNAVLIPEWTIKADYVRKLETKNA  
DVALESTPQNDPEPTQMGNGEGERIHLTPDFAGKRLAEPGAKFVHEMAH  
LRMGVDEYNDEKFLYLSNGRIQAVRCSAGITGNVVKCGGSCYTRKCTFNVTGL

```
YKGECEFLQSRQTEKASIMPAQHDSIVERTCEONHNKKNOKCNLRSTWEI
KDSDEKFTTPTPTPTPTPTSLQIGRIYCLVDKSGSMATGRRLNRNAGQFL
LQTVELSGVMGAVTFDPSAHOVSELIQINSGEDRDTLAKRLPAASGGTSTICSGRSA
FTVIRKRYTPDSEIYLLTDEGNTISGCFNEVKOSGAI IHTVALGPSAODELELSK
MTGQOTASDOVONNGLLDAFGALSGNGAVSOSIOLSEIHLGNSOMNGTVIV
DSYVGKDTLFLITWTPTOPPOLILMDPSGOKOGFVVDNTKAYLOIPGIAKVWTKY
SLQASSQTLTLTVTSRASNNATPTPTVTSKTRKDTSKRPSPLVYANIRQKASPILRA
SVTALIESVNGKTVTLELDNGAGADATKDDVYSIRFTYTDNGRISYKVALGAVN
AARRVVIPOQSGALYIPGMIENDEIQMNPPEIKDQVQKOCFSRSTSGSFLVD
DVNPAPIDPLEPQOITDLKAEIHGSGSLNLTWAPGDYDHGAHKYIIRISTSID
LQDKNESLOVNTTALIPKEANSEVFLFKENITFEENGDTLFLAIQAVDVKLSEI
SNIAVSLFIPQTPPETPSPDETSAPCPNHINSTIGIHILKIMMWIGELQLSIA

BASE COUNT      875 a      623 c      632 g      696 t

ORIGIN
Query Match      100.0%; Score 227; DB 42; Length 2826;
Best Local Similarity 100.0%; Pred. No. 4.7e-125;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTGCGGCGCTCTGGAGAGATTACGACGACGAGAGAGATGATACCCAGCAGAGTGA 60
        |||||||
DB      2027 GTGCGGCGCTCTGGAGAGATTACGACGACGAGAGAGATGATACCCAGCAGAGTGA 2086

QY      61 GCACGTGTACATACCTGGCTGATGTAATGATGAATGCAATGCAATCCACAGACCT 120
        |||||||
DB      2087 GCACGTGTACATACCTGGCTGATGTAATGATGAATGCAATGCAATCCACAGACCT 2146

QY      121 GAAATTAATAGATGATGTTCAACACAGCAAGTGTGTTGACGACATCCCTGGGA 180
        |||||||
DB      2147 GAAATTAATAGATGATGTTCAACACAGCAAGTGTGTTGACGACATCCCTGGGA 2206

QY      181 GGCTCATTTGTGGCTCTGATGTGCCAATGCTCCATACCTGATCT 227
        |||||||
DB      2207 GGCTCATTTGTGGCTCTGATGTGCCAATGCTCCATACCTGATCT 2253

RESULT 4
AF039401      35278 bp      DNA      PRI      15-DEC-1998
LOCUS
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCICL1) gene,
ACCESSION AF039401
NID 94009459
VERSION AF039401.1 GI:4009459
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 35278)
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and
Pauli,B.U.
TITLE Genomic cloning, molecular characterization, and functional
analysis of human CICL1, the first human member of the family of
Ca2+-activated Cl- channel proteins
JOURNAL Genomics 54 (2), 200-214 (1998)
MEDLINE 99047526
REFERENCE 2 (bases 1 to 35278)
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

FEATURES
source
1..35278
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p22-p31"
1..1617
/gene="hCICL1"
1..33522
gene

repeat_region
TATA_signal
exon
mRNA
CDS
exon
/gene="hCICL1"
/ftc_family="L1 transposable element"
/rpl_type="dispersed"
1587..1596
/gene="hCICL1"
/gene="putative"
1618..1708
/gene="hCICL1"
/number=1
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24532..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)
/gene="hCICL1"
/product="calcium-dependent chloride channel-1"
1962..2383
/gene="hCICL1"
/number=2
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24532..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)
/gene="hCICL1"
/codon_start=1
/product="calcium-dependent chloride channel-1"
/protein_id="AAC95429.1"
/db_xref="PID:94009460"
/db_xref="GI:4009460"
/translation="MGPEKSSVFLILHLLEGALSNSLIOLNNGEYIVAVIDPNV
EDERLIQIKRMVQASLIEFATGKRFRFNVAAILIPETKTLADYRPLREYKNA
DVLVAESTPNDPEYTBOMGCGEGRHILTFDPLNGKLAETGPGKAFVEMAH
LRMGVFEDYNDEKFLYSNGRIQAVRCSAGITGNVAKCGGSCYTRKTFNVTGL
YKGECEFLQSRQTEKASIMPAQHDSIVERTCEONHNKKNOKCNLRSTWEI
KDSDEKFTTPTPTPTPTSLQIGRIYCLVDKSGSMATGRRLNRNAGQFL
LQTVELSGVMGAVTFDPSAHOVSELIQINSGEDRDTLAKRLPAASGGTSTICSGRSA
FTVIRKRYTPDSEIYLLTDEGNTISGCFNEVKOSGAI IHTVALGPSAODELELSK
MTGQOTASDOVONNGLLDAFGALSGNGAVSOSIOLSEIHLGNSOMNGTVIV
DSYVGKDTLFLITWTPTOPPOLILMDPSGOKOGFVVDNTKAYLOIPGIAKVWTKY
SVTALIESVNGKTVTLELDNGAGADATKDDGVYSRYFTYTDNGRISYKVALGAVN
SVALAESVNGKTVTLELDNGAGADATKDDGVYSRYFTYTDNGRISYKVALGAVN
AARRVVIPOQSGALYIPGMIENDEIQMNPPEIKDQVQKOCFSRSTSGSFLVD
DVNPAPIDPLEPQOITDLKAEIHGSGSLNLTWAPGDYDHGAHKYIIRISTSID
LQDKNESLOVNTTALIPKEANSEVFLFKENITFEENGDTLFLAIQAVDVKLSEI
SNIAVSLFIPQTPPETPSPDETSAPCPNHINSTIGIHILKIMMWIGELQLSIA"
6667..6807
/gene="hCICL1"
/number=3
6979..7126
/gene="hCICL1"
/number=4
9696..9801
/gene="hCICL1"
/number=5
15454..15631
/gene="hCICL1"
/number=6
18591..18809
/gene="hCICL1"
/number=7
19783..20010
/gene="hCICL1"
/number=8
22253..22427
/gene="hCICL1"
/number=9
24532..24628
/gene="hCICL1"
/number=10
26644..26859
/gene="hCICL1"
/number=11
```





```

OY      75  TGGCTGATGTGAGATGATGA  95
|||||
Db      94913  TGGCTGATGTGAGATGATGA  94893

RESULT  6
AC004067  161326 bp      DNA      PRI      08-NOV-1998
LOCUS
DEFINITION Homo sapiens chromosome 4 clone B366024 map 4q25, complete
sequence.
ACCESSION AC004067
MID       93851208
VERSION   AC004067.1  GI:3851208
KEYWORDS
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 161326)
AUTHORS  Stone,N.E., Keseler,I.M., Schmutz,J.J., Shang,J., Cox,D.R. and
           Myers,R.M.
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 161326)
AUTHORS  Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE     Direct Submission
JOURNAL   Submitted (28-JAN-1998) Department of Genetics, Stanford Human
           Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
           3 (bases 1 to 161326)
REFERENCE Stone,N.E., Keseler,I.M., Schmutz,J.J., Shang,J., Cox,D.R. and
           Myers,R.M.
TITLE     Direct Submission
JOURNAL   Submitted (08-NOV-1998) Department of Genetics, Stanford Human
           Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
           On Nov 9, 1998 this sequence version replaced g1:3694642.
COMMENT   Quality: Phrap Quality >=40 99.8% of Sequence,
           Estimated Total Number of Errors is 0.8
           SIS Content:
           SHGC-19832 G19509
           SHGC-19839 G19514
           SHGC-19837 G19513
           SHGC4-1507 G01867
           SHGC-19823 G19503
           SHGC-51589 G34128
           SHGC-19815 G19498
           SHGC-36185 G30091.

FEATURES
         source
         location/Qualifiers
             1..161326
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="B366024"
                /chromosome="4"
                /map="4q25"
                /clone_lib="Caltech/Research Genetics Whole Genome BAC
                Library"
BASE COUNT  44921 a 33477 c 32754 g 50174 t
ORIGIN
Query Match      8.4%; Score 19; DB 11; Length 161326;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      186  ATTTGTGGCTTGATGTC  204
|||||
Db      58783  ATTTGTGGCTTGATGTC  58801

RESULT  7
M145KDPRA/c  M145KDPRA      2497 bp      DNA
LOCUS
DEFINITION M.1eprae of 45 kd protein.
ACCESSION  Z21952
22-MAY-1995

```

```

NID
VERSION 9311627
KEYWORDS 221952.1 GI:311627
SOURCE 45 kDa protein.
ORGANISM Mycobacterium leprae.
Mycobacterium leprae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.
REFERENCE 1 (bases 1 to 2497)
AUTHORS Rinke de Wit,T.F., Clark-Curtiss,J.E., Abebe,F., Kolk,A.H.,
Janson,A.A., van Agterveld,M. and Thole,J.E.
TITLE A Mycobacterium leprae-specific gene encoding an immunologically
recognized 45 kDa protein
JOURNAL Mol. Microbiol. 10 (4), 829-838 (1993)
MEDLINE 95020554
REFERENCE 2 (bases 1 to 2497)
AUTHORS Rinke de Wit,T.F.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-1993) Rinke de Wit T. F., Armaner Hansen Research
Institute (ARI), Molecular Mycobacteriology, Addis Ababa, Shoa,
Ethiopia, P.O. Box 1005
FEATURES
source location/Qualifiers
1..2497
/organism="Mycobacterium leprae"
/db_xref="taxon:1769"
/clone_lib="M.leprae library (Young et al., Nature,
1985)."
/clone="LL4"
8..13
31..36
/citation=[1]
98..102
113..1339
/citation=[1]
/codon_start=1
/transl_table=1
/evidence=experimental
/product="45 kDa protein"
/protein_id="CAA79950.1"
/db_xref="PID:g311628"
/db_xref="GI:311628"
/db_xref="SWISS-PROT:P007297"
/translacion="MEDPMYISPEVNA*FMSRGGSTPIPMGAFAWISIAEQLEAAQ
EVSOTIVAVPASPAGETSDMLASRVSTFVAMLDGNEAGLIARVLAHVAIATEEAR
AGMPDLTLVGNIIHTMALKAINFGGVSSVVALLEADYDLMVONSTAMTYEDTVY
RETGKMFNEPAPOLYVRVRCMDRDSYVLEFDSSSDSYEDINLDYVQSEHGS
LMSGQSYNTGSGVAVQSELCDSPGCTPGSSQSNLSAPSLTQOGLDSTISSASASI
DTNATSISSSTRASSIMELIVASQYPTETIGRSQAVAYVKMTQSISSSTAVSYDAASKYAVGV
GQAVSVALRVPEHMTATASQPVMTATASVFPAGCSAITTAVSGLPVGTVQPAEEVLTAS
VAGSGTGPGAFENAV"
repeat_region
1548..1574
/citation=[1]
/rpt_unit=1548..1559
/rpt_type=INVERTED
BASE COUNT 533 a 563 c 736 g 663 t
ORIGIN
Query Match 7.9% Score 18; DB 1; Length 2497;
Best Local Similarity 100.0%; Pred. NO. 4.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 145 CACAAGCAAGTGTGTTTC 162
|||||
Db 1920 CACAAGCAAGTGTGTTTC 1903
|||||
RESULT 8
MCL383 35335 bp DNA BCT 27-JUN-1997
LOCUS Mycobacterium leprae cosmid L383.
DEFINITION 291778
ACCESSION 92225939

```

VERSION 297179.1 GI:2225939  
 KEYWORDS bacterioferritin comigratory protein; bcp; repetitive element; RLEP2; RNA methylase; serine-rich antigen; sra; transfer RNA-His; transfer RNA-Lys; transposase pseudogene.  
 SOURCE Mycobacterium leprae.  
 ORGANISM Mycobacterium leprae  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.  
 REFERENCE 1 (bases 1 to 35335)  
 AUTHORS Oliver K. and Harris D.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 35335)  
 AUTHORS Parthill, D., Barrell, B. G. and Rajandream, M. A.  
 JOURNAL Direct Submission  
 Submitted (16-JUN-1997) Mycobacterium leprae sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA. Email: barrell@sanger.ac.uk Cosmids supplied by Dr. Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France Requests for cosmids should be sent to Karin Eiglmeier (kelepasteur.fr)  
 REFERENCE 3 (bases 1 to 35335)  
 AUTHORS Eiglmeier, K., Honore, N., Woods, S. A., Caudron, B. and Cole, S. T.  
 JOURNAL Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae  
 Mol. Microbiol. 7 (2), 197-206 (1993)  
 COMMENT 93188700  
 On Jun 28, 1997 this sequence version replaced gi:2222688.  
 Notes:  
 The Sanger Centre is funded to complete the sequence of M. leprae by the Helser Program for Research in Leprosy and Tuberculosis of The New York Community Trust.  
 Work in Paris is supported by the Helser Trust, the Association Francaise Raoul Follereau and the Groupement de Recherches et des Etudes des Genomes (GIP-GREG).  
 Details of M. leprae sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL, http://www.sanger.ac.uk/projects/)  
 CDS are numbered using the following system eg MLCB33.01c. ML (M. leprae), CB33 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. All CDS over 100 codons have been analysed. Gene prediction is based on positional base preference in codons especially where there is an increase in the observed/expected third position G + C. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid L383 unknown.

FEATURES  
 SOURCE location/qualifiers  
 1..35335  
 /organism="Mycobacterium leprae"  
 /db\_xref="taxon:1769"  
 /clone="cosmid L383"  
 1..496  
 /gene="MLCL383.01"  
 complement(1..9985)  
 /note="overlap with EM.BA:ML015 U00015 Mycobacterium leprae cosmid B1620 from: 2 to: 9984"  
 <1..496  
 CDS  
 stem\_loop  
 gene  
 CDS  
 RBS  
 gene  
 CDS  
 /gene="MLCL383.01"  
 /note="MLCL383.01, unknown, len: 164 aa: highly similar to C-terminus of M. tuberculosis hypothetical protein MTCY07H7B.06 (392 aa), fasta scores: opt: 651 z-score: 605.2 E(): 1.7e-28, 62.2% identity in 164 aa overlap"  
 /codon\_start=2  
 /transl\_table=11  
 /product="hypothetical protein MLCU383.01"  
 /protein\_id="CA009940.1"  
 /db\_xref="PID:e323972"  
 /db\_xref="PID:g2222689"  
 /db\_xref="GI:2222689"  
 /db\_xref="SPTREMBL:Q49722"  
 /translation="SFGGIPGLSGATSGLSQYTGFGVPGLAGSSGLSGSLSTNE NLPISAGVAGIGLIGSLPOLAOLHAASRTGCTSGAGVSAELSTEDPGGQEPVSAOG SCMGSGQSGMGMT PASTKSKDKERKKKTSSEGAAGTDDAENAPIEVSGSGGRALA OHVY"  
 549..552  
 /note="possible RBS upstream of MLCU383.02"  
 554..874  
 /gene="MLCL383.02"  
 554..874  
 /gene="MLCL383.02"  
 554..874  
 /gene="MLCL383.02"  
 /note="MLCL383.02, unknown, len: 106 aa: highly similar to M. tuberculosis hypothetical protein MTCY07H7B.07 (103 aa), fasta scores: opt: 364 z-score: 622.7 E(): 1.8e-29, 60.9% identity in 92 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein MLCU383.02"  
 /protein\_id="CA009941.1"  
 /db\_xref="PID:e323973"  
 /db\_xref="PID:g2222690"  
 /db\_xref="GI:2222690"  
 /db\_xref="SPTREMBL:Q49723"  
 /translation="MRSMDNLTGQSEHLNLSASQENENACASSGSVAAGIANNV TSHGTCQAFNDTLKMTEDNHRTLGSLHRTGIDLARVLVAAMTCDADEICSDIK SAR"  
 887..929  
 /note="hairpin loop with 19bp stem"  
 888..1538  
 /gene="MLCL383.03"  
 888..1538  
 /gene="MLCL383.03"  
 /note="MLCL383.03, unknown, len: 216 aa: highly similar to M. tuberculosis hypothetical protein MTCY07H7B.08 (184 aa), fasta scores: opt: 899 z-score: 1214.3 E(): 0, 71.3% identity in 188 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein MLCU383.03"  
 /protein\_id="CA009942.1"  
 /db\_xref="PID:e323974"  
 /db\_xref="PID:g2222691"  
 /db\_xref="GI:2222691"  
 /db\_xref="SPTREMBL:Q49730"  
 /translation="WGSRRINRRLLPSTPAWQEFRRDYYVFPQNDPDDCDTV DFWGVGAHWITDPIVFASSAPDQDRTVDNTRGQAEDEKQREPAVAFVTNP PRTVSIVLMDRIRHVELSKRTWMSQVASEILVLADLAKRAQSAQVFIIDKL SOLADDEHRYVALLRESVGNWNLPSPEGAABAEVFTRTSDYCPAQDPTNDGW"  
 1845..1991  
 /note="MLCL383.04, transposase pseudogene fragment, len: 48 aa: similar to part of TRAI\_MYCOB\_P35882 transposase for insertion sequence element IS1081 (415 aa), fasta scores: opt: 113 z-score: 207.9 E(): 0.00015, 52.6% identity in 38 aa overlap. Also similar to IS1081 transposase from M. tuberculosis"  
 /codon\_start=1  
 /pseudo  
 /transl\_table=11  
 /db\_xref="PID:e323975"  
 1978..2109  
 /note="MLCL383.05, transposase pseudogene fragment, len:

```

43 aa: similar to part of TRAI MYCB0 P35882 transposase
for insertion sequence element IS1081 (415 aa), fasta
scores: opt: 118 z-score: 214.0 E(): 6.7e-05, 51.28
identity in 43 aa overlap. Also similar to IS1081
transposase from M. tuberculosis"
/codon_start=1
/pseudo
/transl_table=11
/db_xref="PID:e323976"
2110. .2235
/Note="MCL383.06, transposase pseudogene fragment, len:
41 aa: similar to part of TRAI MYCB0 P35882 transposase
for insertion sequence element IS1081 (415 aa), fasta
scores: opt: 136 z-score: 252.5 E(): 4.8e-07, 52.38
identity in 44 aa overlap. Also similar to IS1081
transposase from M. tuberculosis"
/codon_start=1
/pseudo
/transl_table=11
/db_xref="PID:e323977"
2234. .2371
/Note="MCL383.07, transposase pseudogene fragment, len:
45 aa: similar to part of TRAI MYCB0 P35882 transposase
for insertion sequence element IS1081 (415 aa), fasta
scores: opt: 179 z-score: 312.5 E(): 2.2e-10, 69.88
identity in 43 aa overlap. Also similar to IS1081
transposase from M. tuberculosis"
/codon_start=1
/pseudo
/transl_table=11
/db_xref="PID:e323978"
2367. .2447
/Note="MCL383.08, transposase pseudogene fragment, len:
26 aa: similar to part of TRAI MYCB0 P35882 transposase
for insertion sequence element IS1081 (415 aa), fasta
scores: opt: 110 z-score: 277.8 E(): 1.8e-08, 78.38
identity in 23 aa overlap. Also similar to IS1081
transposase from M. tuberculosis"
/codon_start=1
/pseudo
/transl_table=11
/db_xref="PID:e323979"
2434. .2622
/Note="MCL383.09, transposase pseudogene fragment, len:
62 aa: similar to part of TRAI MYCB0 P35882 transposase
for insertion sequence element IS1081 (415 aa), fasta
scores: opt: 197 z-score: 368.4 E(): 1.7e-13, 62.38
identity in 53 aa overlap. Also similar to IS1081
transposase from M. tuberculosis"
/codon_start=1

Query Match 7.9%: Score 18; DB 1; Length 3535;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 CACACAGAGTGTGTTTC 162
|||||
DB 6657 CACACAGAGTGTGTTTC 6640

RESULT 9
LOCUS 000015 42325 bp DNA BCT 04-NOV-1998
DEFINITION Mycobacterium leprae cosmid B1620.
ACCESSION 000015
NID 9466931
VERSION 000015.1 GI:466931
KEYWORDS
SOURCE Mycobacterium leprae.
ORGANISM Mycobacterium leprae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriineae; Mycobacteriaceae;
Mycobacterium.

REFERENCE
AUTHORS 1 (bases 38281 to 41190)
Rinke de Wit,T.F., Bekkele,S., Osland,A., Miko,T.L.,
Hermans,P.W.M., van Soelingen,D., Drijfhout,J.W., Schenningh,R.,
Janson,A.A.M. and Thole,E.R.
Mycobacteria contain two GroEL genes. The second Mycobacterium
leprae GroEL gene is arranged in an operon with GroES
Mol. Microbiol. 6, 1995-2007 (1992)
92374850
TITLE 2 (bases 2759 to 5248)
Rinke de Wit,T.F., Clark-Curtiss,J.E., Abebe,F., Kolik,A.H.,
Janson,A.A., van Agterveld,M. and Thole,J.E.
A Mycobacterium leprae-specific gene encoding an immunologically
recognized 45 kDa protein
Mol. Microbiol. 10 (4), 829-838 (1993)
95020554
JOURNAL MEDLINE 3 (bases 3688 to 5472)
Vega-Lopez,F.
Sequence and immunological characterization of a 42kDa serine-rich
antigen from M. leprae
Infect. Immun. 61, 2145-2153 (1993)
93293328
REFERENCE 4 (bases 1 to 42325)
Smith,D.R.
Unpublished
5 (bases 1 to 42325)
AUTHORS Robison,K.
DIRECT SUBMISSION
Submitted (01-NOV-1993) Department of Genetics, Harvard Medical
School, 200 Longwood Avenue, Boston MA 02115
6 (bases 1 to 42325)
AUTHORS Robison,K.
DIRECT SUBMISSION
Submitted (01-MAR-1994) Department of Genetics, Harvard Medical
School, 200 Longwood Avenue, Boston MA 02115
COMMENT On Mar 31, 1994 this sequence version replaced gi:414221.
This sequence data was produced by the Genome Sequencing Center
located at Collaborative Research Incorporated (1365 Main St.,
Waltham MA, 02159). 617-487-7979. Please contact Doug Smith
(smith@er.cric.com). The annotation should be considered
preliminary and incomplete.
Location/Qualifiers
1. .42325
/organism="Mycobacterium leprae"
/db_xref="taxon:1769"
619. .1808
/Note="trehalose-6-phosphate phosphatase; B1620_F1.1"
/codon_start=1
/transl_table=11
/product="otsp"
/protein_id="AAC43238.1"
/db_xref="PID:9466930"
/db_xref="GI:466950"
/transl_table=11
/translation="MPVTIDPRHSALFDIDAVYDTPLDSTVTLVRQLOGIGVGA
VSTNRNSOGVLTATGLDHPVHAGLGSQVTLVAANRIMAPRCVAVADAA
GITARVSGPALVGLVIGDPTGSRDRLRSAGADTVADIGEVIVRGDRMSLPD
ALQITGLTDDILAROPAVFPFDGTLISIVDDPSARVPQATELQGLATRHCPAIL
SGRDLADVIRKIGVGVIVSGSHGSESTAPGTHQNDAAERTIPILQOATQLRDL
GPIDGVNVEHRKGVAVHYRNVARDRVAVVAATACORNALRVTTREVIELRPDI
DWDKRTLHWYIDRLHAGTOVGSALNPICLDGDTDEDAFDVAVRHVDVGIPIVR
HTEQGNRATLALFTLDSPMHYSEFERLARQLSDIQR"
complement(1979. .2188)
/Note="weak match 42.1 kD protein; B1620_C3_270"
/codon_start=1
/transl_table=11
/product="u1620m"
/protein_id="AAC43237.1"
/db_xref="PID:9466949"
/db_xref="GI:466949"
/transl_table=11
/translation="VCLRLGLAGLKLPIVLSRYGVGRALVVELKGLPNSSIT
LTGVSRRADHLDLENRARADGVSY"
complement(2282. .2473)
/Note="weak match to (gip:x66077); B1620_C3_269"
/codon_start=1

FEATURES
source
CDS
CDS
CDS

```

/transl\_table-11  
/product-"u1620h"  
/protein\_id-"AAC43236.1"  
/db\_xref="PID:q466948"  
/db\_xref="GI:466948"  
/translation-"MDHRSLSCKRQGEHQEPDRCVWMPRIYPGIWDASLIRGRMS  
SRWGSVGTAGCTGFSGS"  
complement(3910..5136)  
/note="45 KDa serine-rich antigen; B1620\_C1\_177"  
/codon\_start-1  
/transl\_table-11  
/product-"ag45"  
/protein\_id-"AAC43220.1"  
/db\_xref="PID:q466932"  
/db\_xref="GI:466932"  
/translation-"MDFMNVSPYEVNAFLMSRPGSTPLMGAAEAMISLABOLMEAAO  
EYSDTIIVAVPASPAGSESDMSLRVSTFVAMLDGNAENAGLIARVLAHVAAYEPEAR  
AGMVLITVIGNIITHTALKAIMEGOVSTVYAALEADYDLMNQNSTAMTYRDTVL  
RETGMENPEPAPQIVSRICMDRRDSVNSFHSSSSDLYESTIDMLYDVAQSEHGS  
DSMSYINTCGVAQSELCDSFPGTSPSSOSNDLSATSLTQGLGDSIISASASL  
LTNNTSSSTASIMPIVAVOYETLGRQVAVERKIQSISSITAVSVVAASKVAVG  
GOAVSVGALRVENMATAQPYMATASVPAGCSAITTAVSGPLEGYTPAEVLTAS  
VAGSGTGPAFENEAV"  
complement(5262..5564)  
/note="weak match 32kd protein (sp.p16645); B1620\_C2\_218"  
/codon\_start-1  
/transl\_table-11  
/product-"u1620k"  
/protein\_id-"AAC43225.1"  
/db\_xref="PID:q466937"  
/db\_xref="GI:466937"  
/translation-"MSFLRYEVGGLMAAGRLERTSESMACNAKLTPTTYKVPYPA  
ADQVSKLSQVFSSTGKQYEGYAAQGVQSRFLVQSLKDAAGDYMSDHYMTNED"  
6340..6816  
/note="transposase; B1620\_F1\_4"  
/codon\_start-1  
/transl\_table-11  
/product-"u1620a"  
/protein\_id-"AAC43241.1"  
/db\_xref="PID:q466953"  
/db\_xref="GI:466953"  
/translation-"MFDLVANRRYGVRLKRLNTAKRAGHDMDRDHDIGVAVRACLAG  
LDGIYDRKIRITREHDQAPGPHRVPDGLVNAKVGSVYESPIISMFRMRCGY  
TAFCDVFSRIRPMPGRTSKTLYVLSLEFQALFIRKYELRFPADFSAPLCGR"  
complement(7286..7606)  
/note="transposase; B1620\_C3\_266"  
/codon\_start-1  
/transl\_table-11  
/product-"u1620b"  
/protein\_id-"AAC43235.1"  
/db\_xref="PID:q466947"  
/db\_xref="GI:466947"  
/translation-"MVNVHALITTDVNAEGYRETLGGCGDHRRRRRGRILEVSAVADRP  
LAFRRGYLGHQRCPRTARRHNGQPPGVWQRCTIYLGHDPHGRLTGFLVGYATP  
LIGY"  
complement(8447..9097)  
/codon\_start-1  
/transl\_table-11  
/product-"B1620\_C3\_264"  
/protein\_id-"AAC43234.1"  
/db\_xref="PID:q466946"  
/db\_xref="GI:466946"  
/translation-"MGSRRIRNRLLMSTFPANOEPRDVVYVFPGRNDRDRCQIV  
DPMGVGAHNTIDPIYGFASSSAPQDRGIDYNTKQAEDEKQKPEVAIFIVNP  
PRIVSVYLMGRLIDHVELSKRTVMSESOVASLILVADLARQKASQAOYTLIDL  
SQLADGEHRVALRESYGTWNLPSPEQAAEAFAEYFATRYSDYCPADOTENDW"  
complement(9111..9431)  
/codon\_start-1  
/transl\_table-11  
/product-"B1620\_C2\_214"  
/protein\_id-"AAC43224.1"  
/db\_xref="PID:q466936"  
/db\_xref="GI:466936"

/translation-"MRSMDNLTVQSEHLSLASQHENEAACASSGVAAGLANAVS  
TSHSGYCAQFNDTLKMYEDAHRLTGESLHTGIDLVAVLVAAMVCDADIEIGSDIK  
SAGS"  
complement(9489..10673)  
/codon\_start-1  
/transl\_table-11  
/product-"B1620\_C2\_213"  
/protein\_id-"AAC43223.1"  
/db\_xref="PID:q466935"  
/db\_xref="GI:466935"  
/translation-"MGSRTIIDPLKAIEMHALLGICVNDGCYVLSISFEKALE  
HLAFAFPDGMVLSADKIVAGQNRKYDIOEALDEKLEIEIHQNASVQTRTGL  
DGAARKALLFVRPAIDLNTYPLGVSVASASIQOACAAAMVSGGLAKLYOIAHT  
AKFVALARLARHLASVADVSDGVAIIKIDYDHLHFIAGALTGKDIYERKIHWF  
EGLFSHWSRLHSFPGIIPGLSGATSGLSOVTGIFGVYGLAGSSGLISGLSTENL  
PSLAGVAGGIGSLPOLADLHAATPOGRSQAQVASEISTOPGQGGCPVVAQSQ  
MGSGSQMGMTPTASTSKDKKKKYSGAAAGTDDAERAFLEVOGGGKRALAOH  
VV"  
13693..13956  
/codon\_start-1  
/transl\_table-11  
/product-"B1620\_F1\_14"  
/protein\_id-"AAC43239.1"  
/db\_xref="PID:q466951"  
/db\_xref="GI:466951"  
/translation-"MSKPLQPECAHIAAVLVGMTNFQAMFLGSPVAIGPAHQAK  
AIRGVGPKTRCALNTSISLPTVTQTTSAVPTREQLRPIALS"  
16793..17830  
/note="D-Ribose periplasmic binding protein; B1620\_F2\_67"  
/codon\_start-1  
/transl\_table-11  
/product-"tbs8"  
/protein\_id-"AAC43242.1"  
/db\_xref="PID:q466954"  
/db\_xref="GI:466954"  
/translation-"MSGTGVPRLVASIMLITGLIPAILPACGFSSEHKRTIGVSP  
TANSPFNATYRTRIDGSRHOLGYINAVSEDEQQLSDYETLHQIDGLIVTQVS  
TSTAFTLRLATQANIPVYVVDYRPGYAPQNNADYVAFVLPNNRKSAGSLAEMLA  
GGSKFLALGMPGNSVAGQKAGLSALTAVGRHLYVOPVYADNEDKGLAAENMLQ  
AHPADGANAIWCFNDKICOGAIKAVYNNANKEFFIGMDLFPQAIATLNGLYTVSL  
GAHMLEGFGELAILYRKIRHQODPAERNVMDLLKVDKSNVAKRYKAYINDTPHYNKS  
TTFPDMTITN"  
18218..19951

CDS

Query Match 7.9%; Score 18; DB 2; Length 42325;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 CACAAGCAAGTGTTC 162  
|||||  
DB 3329 CACAAGCAAGTGTTC 3346

RESULT 10

CCRRATPBE 3762 bp DNA PLN 04-JUN-1996  
DEFINITION C.caldarium rhodoplast atpb, atpe, infb, apcb genes and ori 167.  
ACCESSION X66698  
NID g396522  
VERSION g396522  
KEYWORDS apcb gene; ATP synthase; ATP synthase beta subunit; ATP synthase epsilon subunit; atpb gene; atpe gene; infb gene; Initiation factor

SOURCE  
ORGANISM  
Cyanidium caldarium.  
Cyanidium caldarium.  
Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;  
Porphyridiaceae; Cyanidium.  
1 (bases 1 to 3762)  
AUTHORS  
Kostirzawa,M.  
DIRECT SUBMISSION  
Submitted (25-JUN-1992) M. Kostirzawa, Institut fuer  
Pflanzenphysiologie, JLU Giessen, Heinrich Buff Ring 58, 6300  
Giessen, FRG

REFERENCE 2 (bases 1 to 3762)  
AUTHORS Kostrewa M. and Zetsche K.  
TITLE Organization of plastid-encoded ATPase genes and flanking regions  
including homologs of *infB* and *tsf* in the thermophilic red alga  
*Galdieria sulphuraria*  
JOURNAL Plant Mol. Biol. 23 (1), 67-76 (1993)  
MEDLINE 94033298  
FEATURES

source location/Qualifiers  
1..3762  
/organism="Cyanidium caldarium"  
/strain="14-1-1"  
/db\_xref="taxon:2771"  
/note="rhodoplast"  
complement(1..778)  
/gene="infB"  
complement(<1..778)  
/gene="infB"  
/codon\_start=1  
/product="initiation factor 2"  
/protein\_id="CAA47239.1"  
/db\_xref="PID:G396523"  
/db\_xref="GI:386523"  
/db\_xref="SWISS-PROT:Q08810"  
/translation="MNKINNNLEFNGNDEHEFIEYYNNKKNYSKDLKANKNKTKNKK  
NSNFKSIEKRNRLNNKKNKFLNRRKAYNSNNINCSODIIVENITVNELASK  
IGVATETIKVLFKKGKLLNOLISLNLISQELKINTTIDKKNITVNESNN  
DONINDKVKLRAPVAVLGHVNNKSTLIEKLINDLTKAETGHTTGHGAEFTIG  
FKDKTILIDPGHEAFESIRORVAKLISDILLILIADEGI"  
929..932  
941..1444  
/gene="orf 167"  
941..1444  
/gene="orf 167"  
/codon\_start=1  
/protein\_id="CAA47240.1"  
/db\_xref="PID:G396524"  
/db\_xref="GI:396524"  
/db\_xref="SWISS-PROT:Q08814"  
/translation="MSRSQKNDNFIDKFTIADLILFLPTNNAKAKASYRQKMS  
KQSGEYAEALANYEALNLEDDPYDKSFILYNGILHASGEYKALDYHKALEAN  
NKPQALNNIVYIYHQAQVAKSEINDETAQALFEHAQYKQAKIKAPSYIEAQM  
LSTGRLL"  
1504..2934  
/gene="atpB"  
1504..2934  
/gene="atpB"  
/EC\_number="3.6.1.34"  
/note="beta subunit"  
/codon\_start=1  
/product="H(+)-transporting ATP synthase"  
/protein\_id="CAA47241.1"  
/db\_xref="PID:G396525"  
/db\_xref="GI:396525"  
/db\_xref="SWISS-PROT:Q08807"  
/translation="MSTIAONKGYISQIIGVIEIDPEPNTGLPRTIYNASIKHKNKTI  
TCFYOOGLDNRVRYTAMNSREGLOREIADTGATATVPVCKTCLGIRPVLEPVD  
ELGAIISDRKLPIHRASPLFTOLETKRSIFETGKIVDILAPYKRGKIGLPGGAGV  
KTVLIMELINNIASHGVSVEGGEFRTGNDLVYMKESKYNENKLESEKVALY  
YQGNNEPFGARMRGLTALMAEYFRDNNKODVLFIDNIFRFVQAGSEVALLGKMP  
SAVGOPTLATEMALQDRITSTKESGTSIOAVVPADLIDPAATTFALHDAITV  
LSRGASGIVPADPLDSTSTMLQPOIVGDEHYTAQRKVENLQRYKELQDITAIIG  
LDLSEDRRLTVARAKIERFLISQEPFAVAFYFPGSPKYYSLAKESIDCFKAILNGLD  
DLPEQATYLVGNIDQAIKKADELKDK"  
2924..2927  
/gene="atpB"  
2937..3353  
/gene="atpE"  
2937..3353  
/gene="atpE"  
/EC\_number="3.6.1.34"  
/codon\_start=1  
/product="H(+)-transporting ATP synthase"  
/protein\_id="CAA47242.1"

stem\_loop  
gene  
CDS  
/db\_xref="PID:G396526"  
/db\_xref="GI:396526"  
/db\_xref="SWISS-PROT:Q08808"  
/translation="MGLNIRIADRVIVWNAKAEVILPSTGOLGISHPALITLAL  
DIGVARNITWTSTIVLFGFAVEVDEILILVNGAEASVINLDRANKELIESSSL  
LINEAKTNEKEFEATOKLRKARAVQAAANTLNQSIY"  
3353..3447  
complement(3427..3762)  
/gene="apcb"  
complement(3427..3762)  
/gene="apcb"  
/codon\_start=1  
/protein\_id="CAA47243.1"  
/db\_xref="PID:G396527"  
/db\_xref="GI:396527"  
/db\_xref="SWISS-PROT:P00319"  
/translation="IKDAVAKSLSDSITRRPGGNMYTTRRYAACIRDLDYLRVARY  
SMLAGDSPIIDERVNLGKRYNSLGVPGICATIQSIQAMKEVTSLSVSGEAKKMGIV  
FDYICSGLS"

BASE COUNT 1306 a 522 c 590 g 1344 t  
ORIGIN  
Query Match 7.9%; Score 18; DB 7; Length 3762;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 TTGAGATGATGAAATAC 100  
DB 3142 TTGAGATGATGAAATAC 3159  
RESULT 11  
SPBC660/c 43325 bp DNA PLN 21-DEC-1998  
LOCUS S.pombe chromosome II cosmid c660.  
DEFINITION AL034563  
ACCESSION AL034563  
KEYWORDS NID  
VERSION AL034563.1 GI:4049499  
6-phosphogluconate dehydrogenase decarboxylating; cell wall  
protein; class v pyridoxal phosphate dependent aminotransferase;  
elongation factor g; elongation factor Tu family; fbp1;  
fructose-1,6-bisphosphatase; g beta repeat; glycine-rich protein;  
low-complexity gene-free region; mkl1; mitosis inhibitor protein;  
kinase mkl1; myd like dna-binding protein; neutral trehalase; ntp1;  
ribonucleoprotein; RNA recognition; RNA3' cleavage factor 1b; rpal;  
sbl1; transcription initiation factor 11f beta subunit; WD domain;  
yeast Cf 1b.

SOURCE  
ORGANISM Schizosaccharomyces pombe  
Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
Schizosaccharomycetales; Schizosaccharomycetaceae;  
Schizosaccharomycetes;  
1 (bases 1 to 43325)  
Lyne, M., Rajandream, M.A., Barrell, B.G. and Rieger, M.  
Direct Submission  
Submitted (18-DEC-1998) European Schizosaccharomyces genome  
sequencing project, Sanger Centre, The Wellcome Trust Genome  
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk  
and biotechnologische und molekularbiologische Forschung,  
Angelhofweg 39, D-69225 Mannheim, Germany  
Notes:  
Details of yeast sequencing at the Sanger Centre are available on  
the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/S\\_pombe/](http://www.sanger.ac.uk/Projects/S_pombe/))  
During 1995 to 1996 about 66% of S. pombe chromosome I was  
sequenced by the Sanger Centre. The sequencing of the S. pombe  
genome is now being continued with funding from The European  
Commission. Fourteen European sequencing laboratories, including  
the Sanger Centre, are participating in the project.  
Protein coding regions (CDS) have been predicted with the help of  
computer analysis using the GeneFINDER program in PomBase (an ACEDB

COMMENT

## FEATURES

```

/partial
/gene="SPBC660.01c"
/note="SPBC660.01c, len:463, SIMILARITY: Saccharomyces
cerevisiae, YBR049C, Reb1_YAST, dna-binding protein reb1,
(810 aa), fasta scores: opt: 392, E(): 3e-26, (27.48
identity in 379 aa)"

```

```
misc_feature complement(87..104)
/gene="SPBC660.01c"
/note="ctaatatttaataag, splice branch and acceptor"
misc_feature complement(132..137)
/gene="SPBC660.01c"
/note="gtaaagt, splice donor sequence"
misc_feature complement(412..546)
```

```

/gene="SPBC660.01c"
/note="gtatgt, splice donor sequence"
2618..3929
/gene="SPBC660.02"
join(2618..2738,2785..3929)
CDS
gene

```

```

protein, 613280, swr1, (550 aa), iadscd scores: opt: 1301,
E():0, (51.68 identity in 372 aa)"

```

```

/misc_feature
    /gene="SPBC660.02"
    /note="ctaagacagcag, splice branch and acceptor"
    3212..3322

```

```

gene      /gene="SPBC660.02"
          /note="Pfam match to entry PF00400 G-beta, WD domain,
          G-beta repeat"
          complement(3930..5015)
CDS       /gene="SPBC660.03c"
          complement(101n3930..4184,4288..4431,4491..5015))

```

```

/codon_start=1
/product="transcription initiation factor 11f, beta
subunit"
/protein_id="CA22523.1"
/db_xref="PID:e1360990"
/db_xref="PID:g4049502"
/db_xref="GI:4049502"

```

```

/misc_feature complement(4282..4287)
/gene="SPBC660.03c"
/note="ctaatgaatcatatg, splice branch and acceptor"

```

```
/gene="SPBC660.03c"
/note="gtatgt, splice donor sequence"
complement(6039..7082)
```

```
/product="fructose-1,6-bisphosphatase"
/protein_id="CAA22524.1"
```

```

/db_xref="PID:e1360991"
/db_xref="PID:g4049503"
/db_xref="GI:4049503"
/translation="MKRDLDEIDIDIVLTSFIIQEQRRYNQKRNKNEGKPCIIQENS
GEISLINSIQSFKEFKIANTIRKAEVLVIGISIVSTGDEOKLKDICNDIFITAM
KSNCCGLIYSEEDLIVDNGSVATCPIDGSSNIDGVSIGTIFGIYKLRPS
OGDITDYLBPCKEYVAGYGYMSAHLLTGTGRNVFTIDTIGETILTRMKMP
LOHSITISINCYTAFWDEKIRFTLAHKESTPDKPPKSARTIGSMVADMHTIYIGGL
FAPCSKGNNGKRLLYECFPMALVEQAGIAVNDGDRILDLPVTLHKKSSIMWL
SKHEVEYINFIK"
complement(6066..6833)
/gene="SPBC660.04c"
/note="Pfam match to entry PF00316 FBPAse,
fructose-1-6-bisphosphatase, score 600.50, E-value
9.9e-17"

misc_feature
Query Match 7.9%; Score 18; DB 7; Length 43325;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 AGAATGATGAATACAT 103
|||||
Db 4831 AGAATGATGAATACAT 4814

RESULT 12
SPU33008/c 37596 bp DNA PLN 15-NOV-1995
LOCUS SPU33008 37596 bp DNA PLN 15-NOV-1995
DEFINITION Schizosaccharomyces pombe cosmid 1198.
ACCESSION U33008
NID g1063615
VERSION U33008.1 GI:1063615
KEYWORDS
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Archiascomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces
1 (bases 1 to 37596)
REFERENCE
1 (bases 1 to 37596)
McCombie,W.R.
Sequence analysis of a region of the fission yeast genome
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 37596)
McCombie,W.R.
Direct Submission
TITLE Submitted (02-AUG-1995) W. Richard McCombie, Cold Spring Harbor
AUTHORS Lab, 1 Bungtown Rd, P.O. Box 100, Cold Spring Harbor, NY 11754, USA
JOURNAL Location/Qualifiers
FEATURES
SOURCE
1..37596
/organism="Schizosaccharomyces pombe"
/db_xref="taxon:4896"
/db_xref="cosmid:1198"
/clone_11b="Mizukami, et al., 1993. Cell 73:121-132."
/chromosome="2"
BASE COUNT 12055 a 6803 c 6909 g 11829 t
ORIGIN

Query Match 7.9%; Score 18; DB 7; Length 37596;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 AGAATGATGAATACAT 103
|||||
Db 35354 AGAATGATGAATACAT 35337

RESULT 13
SPU33010/c 85837 bp DNA PLN 15-NOV-1995
LOCUS SPU33010 85837 bp DNA PLN 15-NOV-1995
DEFINITION Schizosaccharomyces pombe cosmid 359, 1198 and 1683.
ACCESSION U33010
NID g1063617
```

```

VERSION U33010.1 GI:1063617
KEYWORDS fission yeast.
SOURCE Schizosaccharomyces pombe
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Archiascomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces
1 (bases 1 to 85837)
REFERENCE
1 (bases 1 to 85837)
McCombie,W.R.
Sequence analysis of a region of the fission yeast genome
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 85837)
McCombie,W.R.
Direct Submission
TITLE Submitted (02-AUG-1995) W. Richard McCombie, Cold Spring Harbor
AUTHORS Lab, 1 Bungtown Rd, P.O. Box 100, Cold Spring Harbor, NY 11754, USA
JOURNAL Location/Qualifiers
FEATURES
SOURCE
1..85837
/organism="Schizosaccharomyces pombe"
/db_xref="taxon:4896"
/clone="cosmids 359 (partial), 1198 and 1683"
/clone_11b="Mizukami, et al., 1993. Cell 73:121-132."
/chromosome="2"
BASE COUNT 27560 a 14781 c 15395 g 28101 t
ORIGIN

Query Match 7.9%; Score 18; DB 7; Length 85837;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 AGAATGATGAATACAT 103
|||||
Db 83595 AGAATGATGAATACAT 83578

RESULT 14
PTU40706/c 3077 bp mRNA PLN 19-NOV-1996
LOCUS PTU40706 3077 bp mRNA PLN 19-NOV-1996
DEFINITION Pachysolen tanophilus aldose reductase gene, complete cds.
ACCESSION U40706
NID g1674391
VERSION U40706.1 GI:1674391
KEYWORDS
SOURCE Pachysolen tanophilus.
ORGANISM Pachysolen tanophilus
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Pachysolen.
1 (bases 1 to 3077)
REFERENCE
1 (bases 1 to 3077)
Boelen,P.L., Hayman,G.T. and Shepherd,H.S.
Sequence and analysis of an aldose (xylose) reductase gene from the
JOURNAL xylose-fermenting yeast Pachysolen tanophilus
MEDLINE Yeast 12 (13), 1367-1375 (1996)
97082506
REFERENCE
2 (bases 1 to 3077)
Boelen,P.L.
Direct Submission
TITLE Submitted (15-NOV-1995) Paul L. Boelen, Flavors Dept., International
AUTHORS Flavors & Fragrances, Corporate R&D, 1515 Highway 36, Union Beach,
JOURNAL NJ 07735, USA
FEATURES
SOURCE
1..3077
/organism="Pachysolen tanophilus"
/isolate="NRRL Y-2460; ATCC 32691"
/db_xref="taxon:4918"
/clone="PBS154"
/note="cDNA library was constructed from mRNA isolated
from a xylose-grown culture"
1031..1967
/EC_number="1.1.1.21"
/codon_start=1
/product="aldose reductase"
/protein_id="AAC49526.1"
```

CDS

```
/db_xref="PID:g1674392"
/db_xref="GI:1674392"
/translation="MTLQYTYTLNNGRRIPALIGCCKWLENADMYTAIREGRLEPDC
ACQYGNKEVEGEENRAIKDGLVKRDLFTSKLMNFWKAKMLKSLSDFINL
YFDLYLHNPISFEPVEEKPFGCGDGKFTYEDVDLIETWRAMENLVDEGLVK
SIGSVSNVSGGLEDLIKAKRKIPASLIQIEHHPLYQOKLAEVYQKGIYVTVGSNRP
LSFLEIGNETAKTQPLXENKTTITIAKKGKTPFOVLAEVYQKGIYVTVGSNRP
LAVNLHDEDFDLKEDEFEETAKLDRLHREFDPTWDXITPTFV"

repeat_region
2614..2901
1016 a 354 c 659 g 1048 t

BASE COUNT
ORIGIN

Query Match 7.98; Score 18; DB 8; Length 3077;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 127 AATAAGATGATGTTCAA 144
1610 AATAAGATGATGTTCAA 1593

RESULT 15
AC000389/c
LOCUS AC000389
DEFINITION AC000389
ACCESSION AC000389
NID g2160128
VERSION AC000389.1 GI:2160128
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 83024)
AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lande,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Brown,K., Cooke,P., Daly,M.J., Forrest,C., Frilip,W.J.,
Gage,D., Geraltger,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,D.,
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
Direct Submission
Submitted (08-APR-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 83024)
AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lande,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Brown,K., Cooke,P., Daly,M.J., Forrest,C., Frilip,W.J.,
Gage,D., Geraltger,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,D.,
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
Direct Submission
Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 4, 1997 this sequence version replaced g1:1929467.
The Staden databases, finishing information, and all
chromatographic files used in the assembly of this clone are
available from our anonymous ftp site.

All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
Location/Qualifiers
1..83024
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="276015"
/clone_lib="Research Genetics/Cal Tech CIRS-HSP-C (plates
195-384)"
/chromosome="17"
repeat_region complement(1586..1648)
repeat_region /rpt_family="AT_rich"
repeat_region complement(1683..1719)
repeat_region /rpt_family="(CAA)n"
repeat_region complement(1722..2011)
repeat_region /rpt_family="AluYo"
repeat_region 2068..2241
repeat_region /rpt_family="MIR"
repeat_region complement(3348..3388)
repeat_region /rpt_family="AT_rich"
repeat_region complement(3684..4538)
repeat_region /rpt_family="L1MA4"
repeat_region complement(3867..4538)
repeat_region /rpt_family="L1PA12"
repeat_region complement(4539..4837)
repeat_region /rpt_family="AluY"
repeat_region complement(4838..5089)
repeat_region /rpt_family="L1PA16"
repeat_region complement(4956..6384)
repeat_region /rpt_family="L1"
repeat_region complement(6384..6458)
repeat_region /rpt_family="L1MA5A"
repeat_region complement(6597..6632)
repeat_region /rpt_family="AT_rich"
repeat_region complement(6696..6737)
repeat_region /rpt_family="L1MC3"
repeat_region complement(6705..6737)
repeat_region /rpt_family="L1ME1"
repeat_region complement(6705..6761)
repeat_region /rpt_family="L1ME3A"
repeat_region 6974..7280
repeat_region /rpt_family="AluSg"
repeat_region complement(7542..7599)
repeat_region /rpt_family="AT_rich"
repeat_region 8322..8397
repeat_region /rpt_family="(TAA)n"
repeat_region complement(9278..9431)
repeat_region /rpt_family="L1HS"
repeat_region 9432..9709
repeat_region /rpt_family="L1PA4"
repeat_region complement(9732..10699)
repeat_region /rpt_family="L1R5"
repeat_region complement(10835..11043)
repeat_region /rpt_family="L1MC3"
repeat_region 12410..12714
repeat_region /rpt_family="L1MA3"
repeat_region 13520..14075
repeat_region /rpt_family="L1MC3"
repeat_region 13823..14085
repeat_region /rpt_family="L1ME2"
repeat_region complement(14336..14455)
repeat_region /rpt_family="AluYo"
repeat_region complement(14456..14590)
repeat_region /rpt_family="L1"
repeat_region complement(14891..14952)
repeat_region /rpt_family="(CA)n"
repeat_region complement(15581..15880)
repeat_region /rpt_family="AluX"
repeat_region 16204..16488
repeat_region /rpt_family="MER42B"
repeat_region 16845..16959
repeat_region /rpt_family="MIR2"
repeat_region 19836..20314
repeat_region /rpt_family="L1"
repeat_region 20411..20605
repeat_region /rpt_family="L1"
repeat_region 20456..21421
```



```

repeat_region /rpt_family="LIMD1"
                21403..21723
repeat_region /rpt_family="LIPB1"
                complement(21731..21765)
                /rpt_family="AT_rich"
repeat_region 21952..22299
                /rpt_family="MLT2FB"
repeat_region complement(22314..22349)
                /rpt_family="(CA)n"
repeat_region 22491..22824
                /rpt_family="MSRA"
repeat_region complement(22838..22963)
                /rpt_family="MIR"
repeat_region complement(23273..23531)
                /rpt_family="AluY"
repeat_region 24170..24255
                /rpt_family="MIR"
                24293..24687
repeat_region /rpt_family="MSRA"
                complement(24718..24902)
                /rpt_family="MIR"
repeat_region complement(26019..26042)
                /rpt_family="AT_rich"
                26280..26579
repeat_region /rpt_family="AluY"
                complement(27026..27048)
                /rpt_family="AT_rich"
repeat_region complement(27617..27923)
                /rpt_family="AluY"
repeat_region complement(28258..28338)
                /rpt_family="MIR2"
                28477..28674
repeat_region /rpt_family="MER20"
                29070..29360
repeat_region /rpt_family="LIMB8"
                complement(29762..29800)
                /rpt_family="MADE1"
repeat_region complement(30161..30521)
                /rpt_family="MLT1A1"
repeat_region complement(32306..32415)
                /rpt_family="(TAAA)n"
repeat_region complement(35059..35579)
                /rpt_family="LIPAZ"
                38096..38263
repeat_region /rpt_family="MIR"
                complement(38948..39019)
                /rpt_family="AT_rich"
repeat_region complement(39210..39232)
                /rpt_family="AT_rich"
repeat_region complement(40840..40891)
                /rpt_family="AT_rich"
                40900..41197
repeat_region /rpt_family="AluSc"
                complement(42137..42216)
                /rpt_family="(TA)n"
repeat_region complement(42452..42515)
                /rpt_family="MIR2"
                42581..42756
repeat_region /rpt_family="MIR"
                complement(43437..43476)
                /rpt_family="AT_rich"
repeat_region complement(43591..43629)
                /rpt_family="AT_rich"
                44199..44482
repeat_region /rpt_family="AluSx"
                44483..44520
repeat_region /rpt_family="(CA)n"
                44520..44549
repeat_region /rpt_family="(GA)n"
                complement(45158..46193)
                /rpt_family="LIPAZ"
                complement(46246..47118)
                /rpt_family="LIPAZ"

```

```

repeat_region complement(46975..51153)
                /rpt_family="LI"
repeat_region complement(52158..53058)
                /rpt_family="LIME1"
repeat_region complement(53431..53692)

Query Match 7.9% Score 18; DB 10; Length 83024;
Best Local Similarity 100.0%; Fred. NO. 4.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 AAGTGTGTTTCACAGAA 169
    |||||
DB 62400 AAGTGTGTTTCACAGAA 62383

```

Search completed: August 6, 1999, 11:41:14  
Job time: 6560 sec

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:18 ; Search time 3075.15 Seconds  
(without alignments)  
145,608 Million cell updates/sec

Title: US-09-049-696-13

Perfect score: 227  
Sequence: 1 GGGCGGCTCTGGAGAGACT.....AATGCTCCATCTGACT 227

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	107.4	47.3	734	49	A1660957	A1660957 wF20d08.x
2	103.2	45.5	469	36	AA611564	AA611564 v085d05.r
3	89.2	39.3	449	38	AA759816	AA759816 v081b01.r
4	78.2	34.4	420	39	A1641868	A1641868 vq39a03.y
5	71	31.3	430	37	AA734088	AA734088 vs19c08.r
6	62.8	27.7	646	37	AA691335	AA691335 vs14h06.r
7	58.2	25.6	479	35	AA541829	AA541829 vj01f01.r
8	55.8	24.6	524	39	AA871630	AA871630 vq39a03.r
9	54.8	24.1	466	29	AA170458	AA170458 ms67c10.r
10	54.8	24.1	468	49	A1640041	A1640041 ms67c10.y
11	47.8	21.1	441	26	W41083	W41083 mc41c10.r1
12	47.8	21.1	470	43	A1197242	A1197242 uc61f01.r
13	36.8	16.2	507	36	AA623217	AA623217 v114d07.r
14	35.2	15.5	554	36	AA623196	AA623196 v114b07.r
15	35	15.4	629	34	AA518112	AA518112 v123c06.r
16	33.4	14.7	472	40	AA914266	AA914266 vq93f08.r
17	32.4	14.3	586	25	N61222	N61222 T9ESTzy09c0
18	31.4	13.8	335	21	T94693	T94693 y631f12.r1
19	31.4	13.8	347	26	W81412	W81412 zd86b11.r1
20	31.4	13.8	467	30	AA242872	AA242872 zr64c01.r
C 21	31.4	13.8	397	30	AA252159	AA252159 zr64c01.s
22	31.4	13.8	455	41	AA96180	AA96180 os26h11.s
C 23	31.4	13.8	626	50	A1683424	A1683424 tw51a12.x
24	30.8	13.6	211	34	AA472982	AA472982 vb31h11.r
25	30.8	13.6	370	37	AA711143	AA711143 vt56d09.r
26	30.8	13.6	426	47	A1480989	A1480989 vf66e07.x
27	30.6	13.5	298	47	A1508480	A1508480 vb31h11.y
28	30.6	13.5	260	20	T08137	T08137 EST06028.in
29	30.4	13.4	480	20	T03421	T03421 IB262 InFan
30	30.4	13.4	256	20	T08136	T08136 EST06027.in
C 31	30.4	13.4	349	20	T23729	T23729 seq574 D4HB
32	30.4	13.4	282	20	T36053	T36053 EST96151.Hu
33	30.4	13.4	442	22	H05672	H05672 y175f10.s1
34	30.4	13.4	400	25	N71293	N71293 z836g05.s1
35	30.4	13.4	392	25	W05353	W05353 z093e01.s1
36	30.4	13.4	474	27	AA030009	AA030009 zk11d02.s
37	30.4	13.4	310	27	AA034907	AA034907 zk25b05.s
38	30.4	13.4	191	27	AA034921	AA034921 zk25d05.s
39	30.4	13.4	396	27	AA035450	AA035450 zk27b03.s
40	30.4	13.4	515	29	AA150278	AA150278 z107g04.s
41	30.4	13.4	426	29	AA179440	AA179440 zp45b06.r
42	30.4	13.4	333	30	AA258305	AA258305 zr60g10.r
43	30.4	13.4	398	34	AA459231	AA459231 aa25d02.r
44	30.4	13.4	380	38	AA743938	AA743938 ob17b01.s
45	30.4	13.4	203	49	A1648690	A1648690 tx64h07.x

#### ALIGNMENTS

RESULT 1  
LOCUS A1660957/c 734 bp mRNA EST 10-MAY-1999  
DEFINITION wF20d08.x1 Soares\_Dieckgraefe\_colon\_NHUC Homo sapiens cDNA clone  
IMAGE:2351151 3' similar to TR:088826 O88826 GOB-5 PROTEIN. ; mRNA  
sequence.  
ACCESSION A1660957  
NID g4764540

VERSION A1660957.1 GI:4764540  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 734)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Tumor gene index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980033.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 477.

FEATURES  
SOURCE  
1..734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="X"  
/clone="IMAGE:2351151"  
/clone.lib="Soares Dieckgraefe.colon\_NHUC"  
/tissue.type="colonic mucosa from 5 ulcerative colitis  
patients"  
/lab.host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified p773 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieckelm.wustl.edu); colonic mucosa represents a range of  
disease involvement from mild cryptitis to severe  
ulceration, fibrosis, and degeneration. Library  
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 200 a 143 c 159 g 228 t 4 others  
ORIGIN

Query Match 47.3%; Score 107.4; DB 49; Length 734;  
Best Local Similarity 84.3%; Pred. No. 3.8e-23;  
Matches 166; Conservative 0; Mismatches 25; Indels 6; Gaps 4;

OY 31 AGACGAGAGTATACCCAGCAGAGTGTACATCTGCTGATGATGAGAT 90  
DB 734 AGACGAGAGTATACCCAGCAGAGTGTACATCTGCTGATGATGAGAT 675  
OY 91 GATGAATATACATGATGATCCACCAAGACCTGAATTAATAGATGATGATCAACAAAG 150  
DB 674 GATG--AATACATATGATGATCCACCAAGACNTGAA--TAATAAGNATGATG--TCAACAAAG 619  
OY 151 CAAGTGTCTTTCAGACAAACATCTCGGGAGGCTATTTGGCTTCGATGATCCCAAT 210  
DB 618 CAAGTGTCTT--CAGCAGACATCTCGGGAGGCTATTTGGCTTCGATGATCCCAAT 561  
OY 211 GCTCCATACCTGATCT 227  
DB 560 GCTCCATACCTGATCT 544

RESULT 2  
LOCUS AA611564 469 bp mRNA EST 01-OCT-1997  
DEFINITION v085d05.r1 Barstead mouse irradiated colon MRLB7 Mus musculus cDNA  
clone IMAGE:1065897 5' similar to TR:G1184066 G1184066

ACCESSION A611564  
NID 92461643  
VERSION A611564.1 GI:2461643  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 469)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
JOURNAL On Sep 12, 1996 this sequence version replaced gi:1400911.  
COMMENT

TITLE  
JOURNAL  
COMMENT

CONTACT: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 457.

FEATURES  
SOURCE  
1..469  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1065897"  
/clone.lib="Barstead mouse irradiated colon MRLB7"  
/dev.stage="8 weeks"  
/lab.host="DH10B  
/note="Vector: p773D-Pac (Pharmacia) with a modified  
polylinker; Site.1: EcoRI; Site.2: NotI; Tissue obtained  
from 8 week old mouse. Colon was harvested 72 hours after  
irradiation with 1400 Gys. 1st strand cDNA was primed  
with a Not I - oligo(dT) primer  
[5'TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'];  
double-stranded cDNA was ligated to Eco RI  
adaptors [AATTCGATCTTG], digested with Not I and cloned  
into the Not I and Eco RI sites of the modified p773  
vector. Library constructed by Bob Barstead.

BASE COUNT 131 a 111 c 123 g 104 t  
ORIGIN

Query Match 45.5%; Score 103.2; DB 36; Length 469;  
Best Local Similarity 73.3%; Pred. No. 6.7e-22;  
Matches 148; Conservative 0; Mismatches 48; Indels 6; Gaps 1;

OY 5 GGGCTCGGAGAGGTTAACCCAGCCAGAGAGTATACCCAGCAGAGTGGAGCAC 64  
DB 162 GGGCTCGGAGAGGTTAACCCAGCCAGAGAGTATACCCAGCAGAGTGGAGCAC 221  
OY 65 TGATACATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 124  
DB 222 TGATACATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 281  
OY 125 TTATAGATGATGTTCAACACAGAGTGTGTTCAACACAGATCCTCGAGAGCT 184  
DB 282 CTA-----GTATAGTTCAAGACAGAGCTGTCTTCAGACAGATCTTCAGGGGGAT 335  
OY 185 CATTTGGCTTCTGATGCTCC 206  
DB 336 CGTTTGGCCACCAATGCTCC 357

RESULT	3
LOCUS	AA759816 449 bp mRNA EST 23-JAN-1998
DEFINITION	vblb01.r1 Barstead mouse irradiated colon MRLRB7 Mus musculus cDNA.
ACCESSION	AA759816
NID	92807610
VERSION	AA759816.1 GI:2807610
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 449)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Districh,N., Dubbque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellendberg,K., Steptoe,W., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R. The WashU-HMI Mouse EST Project Unpublished (1996) On Jan 9, 1998 this sequence version replaced gi:785436.
TITLE	Contact: Marra M/Mouse EST Project
JOURNAL	WashU-HMI Mouse EST Project
COMMENT	Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGJ:635361 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 439. Location/Qualifiers  1..449 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:1177513" /clone_lib="Barstead mouse irradiated colon MRLRB7" /dev_stage="8 weeks" /lab_host="DH10B" /name_vector="pTR3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained from 8 week old mouse. Colon was harvested 72 hours after irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - oligo(dt) primer [5'TGTATCGAATCTGAGTGGAGCGCCGCCCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTGCATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library constructed by Bob Barstead."
FEATURES	
SOURCE	
BASE COUNT	122 a 110 c 115 g 102 t
ORIGIN	
Query Match	39.3%; Score 89.2; DB 38; Length 449;
Best Local Similarity	72.3%; Pred. No.1.3e-17;
Matches 146:	Conservative 0; Mismatches 48; Indels 8; Gaps 2;
OY	5 GGGCTCTGGGAGGATTAACGACGCCAGGAGNGTGTATACCACCAAGAGTGGAGCAC 64
Db	79 GGGCTCTGGGAGGATCTTTCAGACACAGAGAGCAGCACCCTCGAAGAACAAGGCCA 138
OY	65 TGTATACCTCGTGGCATTTGAGATGGAATCAATGGAATCCACCAAGACTGGAAA 124
Db	139 TGTCATATGATGCTCGGATTCGAGATGGTGAAGTAGTAAGCAATGAACCCACACAGCTCTGAAA 198
OY	125 TTAAATAGATGATGTTCAACACAGAAGAGTGCTGTTTCAGCAGAATCTTCGGAGGCT 184

```

DB      199 CTA-----GTTATGTCTAGACGAACGACTGT--CACAGGACATCTTCAGGGGAT 250
OY      185 CATTTTGCGCTTGCATGTCCC 206
Db      251 CGTTGTGGCACCACCAATGTCC 272

RESULT   4
LOCUS     A1641868
DEFINITION vq39a03.y1 Barstead bowel MPRB9 Mus musculus cDNA clone
IMAGE:1096588 5' similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA
sequence.
ACCESSION A1641868
NID        94720343
VERSION    A1641868.1 GI:4720343
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 420)
AUTHORS    Marra,M., Hallier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
           Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y.,
           Ritzer,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
           Waterston,R. and Wilson,R.
TITLE       The Mashu-NCI Mouse EST Project 1999
JOURNAL    Unpublished (1999)
COMMENT    On Mar 10, 1998 this sequence version replaced gi:2948637.

Contact: Marra M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:602820

This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40RP from GIBCO
High quality sequence stop: 389.
Location/Qualifiers
1..420
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1096588"
/clone_id="Barstead bowel MPRB9"
/tissue-type="bowel"
/lab_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: pUT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: EcoRI; Site.2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACGAACTCTGAAGCGGAGCGGCCGCTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGAATCTTGG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pUT73 vector.
Source irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob
Barstead."

BASE COUNT      117 a          98 c          113 g          92 t
ORIGIN
Query Match      34.4% Score 78.2; DB 49; Length 420;
Best Local Similarity 62.1%; Pred. NO. 3.2e-14;
Matches 141; Conservative 0; Mismatches 83; Indels 3; Gaps 1;

```















This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:03:46 ; Search time 921.96 Seconds

(without alignments)  
67.300 Million cell updates/sec

Title: US-09-049-696-14

Perfect score: 248  
Sequence: 1 ACCGTGAGCGCGAATTCAC.....TTGAAAATGGCAGATCTT 248

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	248	100.0	878	1 T45884	Human colon specif
2	248	100.0	878	1 V16672	Poly nucleotide seq
3	33	13.3	1984	1 T13030	Cotton fibre-speci
4	33	13.3	1985	1 T130250	Cotton fibre clone
5	33	13.3	1985	1 T70036	Cotton H6 gene and
6	30.8	12.4	1570	1 V07180	Ehrlichia canis VS
7	30.8	12.4	864	1 X34760	DNA encoding P30a
8	30.2	12.2	3653	1 X20280	Borrelia burgdorfe
9	29.8	12.0	273	1 T20583	Human gene signatu
10	29.8	12.0	3496	1 V35254	Candida albicans C
11	29.8	12.0	110000	1 V21209_01	Continuation (2 of
12	29.6	11.9	18177	1 N90490	DNA of human retin
13	29.6	11.9	6313	1 V74512	Staphylococcus aur
14	29.6	11.9	110000	1 X20248_06	Continuation (7 of
15	29.6	11.9	111309	1 X04502	Borrelia burgdorfe
16	29.6	11.9	18303	1 X20250	Human retinoblasto
17	29.2	11.8	110000	1 T58840_1	Continuation (2 of
18	29.2	11.7	1730	1 O46580	5-HT1F receptor cl
19	29	11.7	1554	1 Q72272	Human NR77 seroton
20	29	11.7	11283	1 T84841	Human breast and o
21	29	11.7	562	1 X13725	Enterococcus faeca
22	28.8	11.6	804	1 T67526	H. pylori secreted
23	28.8	11.6	825	1 T67960	H. pylori secreted
24	28.8	11.6	4289	1 X13247	Enterococcus faeca
25	28.8	11.6	951	1 X20046	Enterococcus faeca
26	28.8	11.6	871	1 X20047	Enterococcus faeca
27	28.6	11.5	2881	1 V74565	Staphylococcus aur
28	28.4	11.5	4008	1 O64827	Leukotoxin ApplIIA
29	28.4	11.5	800	1 T28546	E. faecalis detect
30	28.4	11.5	7721	1 T73220	ApplIIABCD gene. P
31	28.4	11.5	29729	1 X13175	Enterococcus faeca
32	28.2	11.4	1742	1 N00004	Sequence of Influe
33	28.2	11.4	2730	1 T32583	Alfalfa isoflavone
34	28.2	11.4	676	1 T58731	LsrDNA sequence fr
35	28.2	11.4	676	1 T58732	LsrDNA sequence fr
36	28.2	11.4	676	1 T58733	LsrDNA sequence fr
37	28.2	11.4	1761	1 V44900	Petunia flower bud
38	28.2	11.4	5277	1 X13199	Enterococcus faeca
39	28.2	11.4	894	1 X20462	Human secreted pro
40	28	11.3	4704	1 O05264	Sequence encoding
41	28	11.3	971	1 O23358	Anther-specific pr
42	28	11.3	1659	1 O74818	Firefly luciferase
43	28	11.3	2063	1 O74819	Firefly luciferase

C 44 28 11.3 2063 1 O74819  
C 45 28 11.3 312 1 T26084

Firefly luciferase  
Human gene signatu

## ALIGNMENTS

RESULT 1

T45884 standard; cDNA: 878 BP.

AC T45884;

DE Human colon specific gene CSG5 cDNA partial clone.

KW Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 2..691 /tag= a

PT W09639419-A1.

PD 12-DEC-1996.

PE 06-JUN-1995; U07289.

PR 06-JUN-1995; WO-U07289.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Yu G,

DR WPI: 97-043054/04.

DR P-PSDB: W06548.

PT Human colon specific genes and their expression products - detection

of which, in non-colon tissue samples, can be used as indication of

colon cancer metastasis

PS Claim 1: Fig 5; 60pp; English.

CC 13 cDNA clones (743860-92), most of them partial clones, correspond

CC to human colon specific genes, designated CSG1, CSG2, etc., that

CC are primarily expressed in tissues derived from the colon. CSG7

CC and CSG10 show reduced expression in colon cancer cells as compared

CC to that in normal cells; the remaining genes are overexpressed in

CC colon cancer. The partial cDNA sequences can be used to isolate

CC full-length clones and genomic clones including the complete gene.

CC CSG nucleic acids can be used to produce CSG polypeptides (see also

CC W06545-53) in transformed host cells, as probes to detect disorders

CC of the colon, partic. colon cancer and colon cancer metastasis, and

CC in gene therapy.

SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 100.0%; Score 248; DB 1; Length 878;

Best local similarity 100.0%; Pred. No. 1.7e-68;

Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AC V16672:  
 DT 22-JUN-1998 (first entry)  
 DE Polynucleotide sequence of a colon-specific gene.  
 KW Colon-specific gene; probe; detection; expression; human;  
 OS Homo sapiens.  
 FT Key location/Qualifiers  
 FT CDS 2..685  
 FT /note= "a  
 FT /note= "no stop codon given"  
 PN US5733748-A.  
 PD 31-MAR-1998.  
 PE 06-JUN-1995: 469667.  
 PR 06-JUN-1995: US-469667.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen C, Yu G.  
 DR MPI: 98-229823/20.  
 DR P-PSDB: W46879.  
 PT Colon-specific nucleic acids - useful as probes for detecting colon  
 PT cancer micrometastases  
 PS Claim 15: Fig 5A-B; 51bp; English.  
 CC V16672-81 represent polynucleotide sequences of partial or full-length  
 CC cDNA clones of colon-specific genes. The polynucleotides can be used  
 CC as probes to detect expression of the corresponding human genes, e.g. in  
 CC diagnostic assays for detecting micrometastases of colon cancer.  
 CC Recombinant cells containing the polynucleotides can be used to  
 CC produce the polypeptides, in order that antibodies can be raised and  
 CC used in further screening or diagnostics.  
 CC Sequence 878 BP: 257 A; 179 C; 188 G; 241 T;  
 SQ

Query Match 100.0%; Score 248; DB 1; Length 878;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-68;

Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTGAGGCGGAAATTCACGGGGCAGTCTCATTTATCTGACCTGACAGCTCCTGGG 60  
 DB 306 ACCTGAGGCGGAAATTCACGGGGCAGTCTCATTTATCTGACCTGACAGCTCCTGGG 365  
 QY 61 ATGATTATGACATGGAACAGCTCACAAGTATATTCATTCGATTAAGTACAGTATCTTG 120  
 DB 366 ATGATTATGACATGGAACAGCTCACAAGTATATTCATTCGATTAAGTACAGTATCTTG 425  
 QY 121 ATCTGAGAGACAGTTCATGATCTCTCAATGATATCTGCTCTCATCCCAAGG 180  
 DB 426 ATCTGAGAGACAGTTCATGATCTCTCAATGATATCTCTCTCATCCCAAGG 485  
 QY 181 AAGCCAACTCGAGAGAGCTTTTGTGTTAAACGAAACATTACTTTGAAATGCGA 240  
 DB 486 AAGCCAACTCGAGAGAGCTTTTGTGTTAAACGAAACATTACTTTGAAATGCGA 545  
 QY 241 CAGATCTT 248  
 DB 546 CAGATCTT 553

RESULT 3  
 T13030/c  
 ID T13030 standard; DNA: 1984 BP.  
 AC T13030;  
 DT 27-MAY-1996 (first entry)  
 DE Cotton fibre-specific H6 gene and promoter.  
 KW Cotton; fibre; promoter; transgenic plant; crop improvement; ds.  
 OS Gossypium barbadense strain Sea Island.  
 FH Key Location/Qualifiers  
 FT promoter 1..250  
 FT /tag= a  
 FT /note= "contains promoter sequence (claimed)"  
 FT 5'utr 251..320  
 FT /tag= b  
 FT /note= "5' non-coding sequence"  
 FT cds 321..1547  
 FT /tag= c  
 FT /note= "the CDS includes a 583 bp intron  
 FT (location unspecified)"

FT 3'utr 1548..1984  
 FT /tag= d  
 PN US5495070-A.  
 PD 27-FEB-1996.  
 PE 04-OCT-1988: 253243.  
 PR 04-OCT-1988: US-253243.  
 PR 21-NOV-1990: US-617239.  
 PR 18-MAY-1992: US-885970.  
 PA (CETU) AGRACETUS INC.  
 PI John M.  
 DR MPI: 96-139095/14.  
 PT New isolated fibre-specific promoters - used for introducing  
 PT altered fibre-specific characteristics into plants, partic. cotton.  
 PS Claim 1: Column 59-62; 48pp; English.  
 CC The cotton var. Sea Island H6 gene and promoter (T13030) were  
 CC identified in genomic clone EMBL-S1-H6-4 isolated by screening a  
 CC genomic library prep. in lambda EMBL with cotton fibre-specific  
 CC cDNA clone H6 (see T13034). H6 and other (see T13025-29, T13031-32  
 CC and T13052-53) fibre-specific promoters can be utilized in the  
 CC construction of vectors used for the introduction of altered fibre-  
 CC specific characteristics into plants, partic. cotton. They can be  
 CC used to modulate the synthesis of fibre proteins or to introduce  
 CC non-fibre proteins into fibre in a tissue-specific manner  
 CC Sequence 1984 BP: 565 A; 470 C; 281 G; 668 T;  
 SQ

Query Match 13.3%; Score 33; DB 1; Length 1984;  
 Best Local Similarity 53.5%; Pred. No. 0.36;

Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 90 TATATCATTCGAATTAAGACAGATTTCTGATCTCAGAGCAAGTCAATCTCTT 149  
 DB 1817 TATATGAATTCGAATTAAGACAGATTTCTGATCTCAGAGCAAGTCAATCTCTT 1758  
 QY 150 CAATGATTAATCTGCTCTCATCCCAAGAGCAAGTCTGAGGAGCTTTGTT 209  
 DB 1757 CAATGATTAATCTGCTCTCATCCCAAGAGCAAGTCTGAGGAGCTTTGTT 209  
 QY 210 AACCAGAA 218  
 DB 1697 AACAATATTA 1689

RESULT 4

T30250/c  
 ID T30250 standard; DNA: 1985 BP.

AC T30250;  
 DT 16-DEC-1996 (first entry)  
 DE Cotton fibre clone SK516-H6-R1-derived promoter contg. sequence.  
 KW Cotton fibre; promoter; differential screening; leaf; ovule; root;  
 KW flower; PCR; polymerase chain reaction; homology; transgenic plant; ds.  
 OS Gossypium barbadense.  
 FH Key Location/Qualifiers  
 FT promoter 1..250  
 FT /tag= a  
 FT /note= "contains promoter sequence (claimed)"  
 FT misc-feature 251..321  
 FT /tag= b  
 FT /note= "5' non-coding sequence"  
 FT cds 322..1547  
 FT /tag= c  
 FT /note= "contains a 583 bp intron"  
 FT misc-feature 1548..1984  
 FT /tag= d  
 FT /note= "3' non-coding sequence"  
 PN US5521078-A.  
 PD 28-MAY-1996.  
 PE 04-OCT-1988: 253243.  
 PR 04-OCT-1988: US-253243.  
 PR 21-NOV-1990: US-617239.  
 PR 18-MAY-1992: US-885970.  
 PR 19-OCT-1994: US-298687.  
 PA (CETU) AGRACETUS INC.

PI John M.; WPI: 96-26794/27.

PT Isolation of fibre-specific cotton promoter sequences - using

PT selected DNA probes to screen genomic DNA fragments, for production

PT of cotton fibres with improved characteristics

PS Claim 1; Column 59-62; 46pp; English.

CC Cotton fibre cell-specific promoter sequences were isolated by

CC differential screening of a cotton plant cDNA library. of 4788 clones

CC from a 10 day cell library screened with leaf cDNAs, 800 clones not

CC present in the leaf were isolated. These were screened with cDNAs from

CC ovule, root and flower mRNAs and resulted in 79 clones isolated. PCR

CC analysis was then used to remove cross-hybridising clones. This resulted

CC in the isolation of 18 cDNA clones specifically expressed in cotton fibre

CC cells (T30242-4 and T30253-67). These cDNAs were then used to screen for

CC homologous genomic sequences (T30245-53 and T30268) in order to obtain

CC the corresp. promoter sequences.

CC This genomic clone is derived from the 1.9 kb SalI-EcoRI fragment of the

CC sub-clone designated SKSIE6-H6-RI. This forms part of the genomic clone

CC designated as EMBL-SI-H6-4 which contains a 13 kb MboI insert.

CC The promoters isolated from the fibre cell-specific clones can be used to

CC generate transgenic cotton plants and lines producing fibres having

CC altered quantity and quality.

SQ Sequence 1985 BP; 565 A; 470 C; 282 G; 668 T;

Query Match 13.3%; Score 33; DB 1; Length 1985;

Best Local Similarity 53.5%; Pred. No. 0.36;

Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0

OY 90 TATATCATTCGATATAGTACGATATTTGTGATCGAGACAGTTCATGATCTT 149

Db 1818 TATATGATCGATATATATCATCAACCAACATATTTCCAAACAAATTAACATTAATTCCTT 1759

OY 150 CAAGTGATATCTACTGCTCTCATCCCAAGGAGCCCAACTCTGAGAGAACTTTTGT 209

Db 1758 CAATATCAAAAAAAGAGTGCTCTCAAAAGSTAAAAAATATATATATGTCCTCAAGAAAT 1699

OY 210 AAACCAAGAA 218

Db 1698 ACAATATATA 1690

RESULT 5

T70036/c

ID T70036 standard; DNA; 1985 BP.

AC T70036;

DE 20-AUG-1997 (first entry)

DT Cotton H6 gene and fibre-specific promoter from clone SIH6.

KW cotton; E6; fibre; promoter; transgenic plant; truncated;

KW heterologous gene expression; ds.

OS Gossypium barbadense strain Sea Island.

PN US5620882-A.

PD 15-APR-1997.

PF 04-OCT-1988; 253243.

PR 04-OCT-1988; US-253243.

PR 21-NOV-1990; US-617239.

PR 18-MAY-1992; US-885970.

PR 19-OCT-1994; US-298829.

PA (CENTU ) AGRACETUS INC.

PI John M.

DR WPI: 97-235185/21.

PT DNA constructs contg. truncated promoter sequence - for

PT fibre-specific gene expression in cotton plants

PS Claim 1; Column 61-64; 48pp; English.

CC T70031-38 are genomic DNA clones containing cotton fibre-specific

CC promoters. Claimed DNA constructs comprise a truncated promoter sequence

CC (from one of T70031-38) that promotes preferential gene expression in

CC plant fibre cells, a protein coding sequence not naturally associated

CC with the promoter sequence and a 3' termination sequence. The DNA

CC constructs are useful for expressing foreign genes in fibre-producing

CC plants, esp. to produce transgenic cotton plants with varied cotton fibre

CC characteristics and quality. The present sequence comprises a fragment of

CC pSIH6-4RI (an EcoRI fragment isolated from pKSIH6-4 which hybridises to

[illegible]

CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,  
 CC Anaplasma and Coxiella species. The nucleic acids are also useful  
 CC as probes to identify related sequences, e.g. for identification of  
 CC organisms and for diagnosing infection. Use of nucleic acid  
 CC vaccines avoids the problem of protein purification associated with  
 CC protein-based vaccines. The nucleic acid does not replicate in the  
 CC host but remains episomal and capable of expressing polypeptide for  
 CC at least 19 mth.  
 CC Sequence 1570 BP; 565 A; 223 C; 251 G; 531 T;

Query Match 12.4%; Score 30.8; DB 1; Length 1570;  
 Best Local Similarity 55.7%; Pred. No. 1.6;

Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 56 TGGGATGATATGACCAAGCAAGCTCACAAGTATATCATTCGAATAGTACAAATAT 115  
 DB 472 TAGCGAGATGCTGACACTGCAAAAGCTGATAGTTGCTCTGAAAATGAAGTTT 531  
 QY 116 TCTGATCTCAGAGCAAGTCAATGATCTCTCAAGCAAACT 161  
 DB 532 ACTTGACGCTCTCATTTATGTTAAAGCATGTTATGACATTAACAAC 577

RESULT 7

ID X34760 standard; DNA; 864 BP.

AC X34760.

DT 05-JUL-1999 (first entry)

DE DNA encoding P30a protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 detection; dog; ss.

OS Ehrlichia canis.

PN W0913720-A1.

PD 25-MAR-1999.

PF 18-SEP-1998; U19600.

PR 19-SEP-1997; US-059353.

PA (OHIS) UNIV OHIO STATE.

PI Ohashi N, Rikihisa Y.

DR WPI: 99-254290/21.

P-PCDB: Y06960.

PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis

PS Disclosure; Fig 20A; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown

CC in Y06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in Y06959-970. The proteins and genes are used

CC to detect E. chaffeensis in patients and E. canis in dogs.

CC Sequence 864 BP; 317 A; 131 C; 135 G; 281 T;

Query Match 12.4%; Score 30.8; DB 1; Length 864;  
 Best Local Similarity 55.7%; Pred. No. 1.3;

Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 56 TGGGATGATATGACCAAGCAAGCTCACAAGTATATCATTCGAATAGTACAAATAT 115  
 DB 462 TAGCGAGATGCTGACACTGCAAAAGCTGATAGTTGCTCTGAAAATGAAGTTT 521  
 QY 116 TCTGATCTCAGAGCAAGTCAATGATCTCTCAAGCAAACT 161  
 DB 522 ACTTGACGCTCTCATTTATGTTAAAGCATGTTATGACATTAACAAC 567

RESULT 8

ID X20280 standard; DNA; 3653 BP.

AC X20280.

DT 04-MAY-1999 (first entry)

DE Borrelia burgdorferi polynucleotide sequence #33.

KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;

KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 infection; diagnosis; characterisation; detection; ds.

OS Borrelia burgdorferi.

PN W0958943-A1.

PD 30-DEC-1998.

PF 18-JUN-1998; U12764.

PR 03-SEP-1997; US-057483.

PR 20-JUN-1997; US-050359.

PR 22-JUL-1997; US-053344.

PR 22-JUL-1997; US-053377.

PA (HUMA-) HUMAN GENOME SCI INC.

PI (MEDI-) MEDIMUNE INC.

PI Clayton R, Dougherty BA, Fraser C, Iathigra R, Smith HO,

PI White OR;

DR WPI: 99-081217/07.

PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease

PS Claim 1; Page 994-996; 112pp; English.

CC X20248 to X20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for

CC the detection, diagnosis, characterisation, prevention and therapy of

CC Bb infections, e.g. Lyme disease. They can also be used for the

CC production of biosynthetic products, e.g. enzymes. Borrelia belongs

CC to a family of motile, spiral-shaped bacteria called Spirochetes.

CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and

CC endemic relapsing fever, and Lyme borreliosis, more commonly known as

CC Lyme disease.  
 CC Sequence 3653 BP; 1587 A; 348 C; 690 G; 1027 T;

Query Match 12.2%; Score 30.2; DB 1; Length 3653;  
 Best Local Similarity 51.1%; Pred. No. 3.3;

Matches 71; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 59 GGATATATGACATGACAGCTCACAAGTATATCATTCGAATAGTACAAATATCT 118  
 DB 1871 GGAGATATATGACAGCTCACAAGTATGCAAAAGCTTAAAGAAATGATGCTAGCA 1930  
 QY 119 TGATCTCAGAGCAAGTCAATGATCTCTCAAGTATATCTGATCTCATCCCAA 178  
 DB 1931 TAATTTAGAACCAAAATTAATGAGGTAAAGCATATATTTGATCTAGAACGAC 1990  
 QY 179 GGAAGCCAACTCTGAGGA 197  
 DB 1991 TGAACCCCAATTAAGAA 2009

RESULT 9

ID T20583 standard; cDNA to mRNA; 273 BP.

AC T20583.

DT 01-AUG-1996 (first entry)

DE Human gene signature HUMG501798.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN W09514772-A1.

PD 01-JUN-1995.

PF 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

DR WPI: 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1; Page 677; 2245pp; Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences



CC given in T19001-726837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 273 BP: 78 A; 77 C; 46 G; 63 T;

Query Match 12.0%; Score 29.8; DB 1; Length 273;  
Best Local Similarity 64.2%; Pred. No. 1.8;  
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 18 CACGGGGGAGTCTCATATCTGACTTGACAGCTCTGGGATGATTATGACATGGA 77  
DB 165 CACGCTGCACCTCATATATCAACTGGCTGCTCTAGATGACATGACATGGA 224

QY 78 ACAGCTC 84  
DB 225 ACACATC 231

RESULT 10  
V32554  
ID V32554 standard; RNA; 3496 BP.

AC 13-OCT-1998 (first entry)  
DE Candida albicans CAC1A4 gene.  
KW CAC1A4: protein kinase; Ste20p family; screening; virulence;  
KW hyphal formation; pathogenic fungi; inhibitor; inflammation;  
KW antimycotic; ss.  
OS Candida albicans.  
FH Key Location/Qualifiers  
FT 432..3347  
FT CDS  
FT /tag= a  
FT /product= CAC1A4 protein

PN W09818927-A1.  
PF 07-MAY-1998.  
PF 29-OCT-1997: CA0809.  
PR 30-OCT-1996: US-029458.  
PR (CANADA) NAT RES COUNCIL CANADA.  
PI Leberer E, Thomas DY;  
PI WPI: 98-27222/24.  
DR P-PSDB; W48896.

PT In vitro screening test for agents that inhibit Candida genes  
PT involved in virulence - and transition to hyphal form, potentially  
PT useful as antimycotic agents  
PS Disclosure: Fig 7; 79pp: English.  
CC The sequence is that encoding the CAC1A4 protein which can be used  
CC in the development of an in vitro screening test for compounds  
CC that inhibit biological activity of the protein and a system for  
CC measuring its activity. The protein is involved in virulence and  
CC hyphal formation. Inhibitors are potentially useful for rendering  
CC pathogenic fungi (any species in which hyphal induction by kinase  
CC occurs) avirulent and/or to treat inflammation. The coding sequence  
CC can be used as source of probes for detecting C. albicans in  
CC amplification or hybridisation assays, also to identify and  
CC clone homologous genes from other fungi.  
SQ Sequence 3496 BP: 1280 A; 598 C; 525 G; 1093 T;

Query Match 12.0%; Score 29.8; DB 1; Length 3496;  
Best Local Similarity 51.9%; Pred. No. 4.3;  
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 43 CTTGGACAGCTCTGGGATGATTATGACCATGGAACAGCTCAAGTATATCATTCGAA 102

DB 505 CTGGCTCAGGTTCTGCTCAGGTTCTGTTCTGCTTCTGCACTTTGGCTAATTTGTTA 564  
QY 103 TAAGTCAAGTATTTCTTGATCTCAGACAGTCAATGATCTCTTCAAGGAAATACTA 162  
DB 565 CCAAGTCTAATAGTCTGTGGCGTAACAGCAATCAACAACTATCAATTAATATA 624  
QY 163 CTGCTCTCA 171  
DB 625 ATCTAGCA 633

RESULT 11  
V21209\_01/c  
Continuation (2 of 17) of V21209 from base 100001 (Methanococcus jannaschii circular  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209  
WP Fragment Name Begin End  
WP V21209\_00 1 110000  
WP V21209\_01 100001 210000  
WP V21209\_02 200001 310000  
WP V21209\_03 300001 410000  
WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000  
WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000  
WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976

Query Match 12.0%; Score 29.8; DB 1; Length 110000;  
Best Local Similarity 54.0%; Pred. No. 14;  
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 27 AGTCTCATATCTGACTGACAGCTCCTGGGATGATTATGACCATGGAACAGCTCAC 86  
DB 22602 AGTGTATATTAACAGATATGATGCTCTTGGAATTAATAGAGTAATATTA 22543

QY 87 AAGTATATCATTCGAAATTAAGTACAACTATTCGATCTCAGAGCAAGTTCAA 139  
DB 22542 AAGGTATATTAATAGAGAAATATAATATGTTGTCACATATTAATATAA 22490

RESULT 12  
N90490/c  
ID N90490 standard; DNA; 18177 BP.  
AC N90490;  
DT 30-OCT-1989 (first entry)  
DE DNA of human retinoblastoma gene exons with flanking regions.  
KW Human; retinoblastoma gene; flanking regions; CDNA;  
OS Homo sapiens (human).  
FH Key Location/Qualifiers  
FH exon 1 533..669  
FH intron 1 1008/1009  
FH exon 2 1318..1444  
FH intron 2 1712/1713  
FH exon 3 1995..2110  
FH intron 3 2363/2364  
FH exon 4 2558..2678  
FH intron 4 2933/2934  
FH exon 5 3227..3265  
FH intron 5 3549/3550  
FH exon 6 3695..3762  
FH intron 6 3983/3984  
FH exon 7 4285..4395

FT intron (1.8 kb) 4639/4640  
FT exon 8 5248/5249  
FT intron (1.8 kb) 5480/.5561  
FT exon 9 5734/5735  
FT intron (1.9 kb) 6000/.6109  
FT exon 10 6381/6382  
FT intron (1.0 kb) 6690/.6768  
FT exon 11 7061/7062  
FT intron (3.2 kb) 7375/.7462  
FT exon 12 7765/7766  
FT intron (3.1 kb) 7996/.8112  
FT exon 13 8349/8350  
FT intron (1.8 kb) 8680/.8736  
FT exon 14 9140/.9171  
FT exon 15 9252/.9328  
FT exon 16 9629/9630  
FT intron (1.0 kb) 9879/.10075  
FT exon 17 10322/10323  
FT intron (70.0 kb) 10564/.10682  
FT exon 18 10911/10912  
FT intron (3.0 kb) 11225/.11370  
FT exon 19 11561/11562  
FT intron (2.8 kb) 11809/.11955  
FT exon 20 12278/12279  
FT intron (4.0 kb) 12538/.12642  
FT exon 21 12948/12949  
FT intron (0.2 kb) 13204/.13317  
FT exon 22 13411/.13574  
FT intron (7.4 kb) 13883/13884  
FT exon 24 14209/.14239  
FT intron (2.8 kb) 14641/14642  
FT exon 25 14834/.14977  
FT exon 26 15488/.15538  
FT intron (1.7 kb) 15952/15953  
FT exon 27 16143/.16213  
FT polyA signal 18010

W08906703-A.  
PD 27-JUL-1989.  
PF 23-JAN-1989: U00293.  
PR 21-JAN-1988: US-146525.  
PA (DRYJ) Dryja T P.  
PI Dryja T P; Friend S; Yandell D W.  
DR WPI: 89-233856/32.  
P-PSDB: P90599.  
PT New human retino-blastoma gene and polypeptide(s)  
PT - used for screening individuals for defective retino-blastoma  
PT gene and treating these patients.  
PS Disclosure: fig 6; 71 pp; English.  
CC DNA encoding human retinoblastoma (RB) gene exons and flanking regions.  
CC The cDNA (see P90599 and N90490) can be used to screen individuals  
CC for the presence of the mutated RB gene. The RB polypeptide  
CC can prevent retinoblastoma formation, and corresp. antibodies  
CC can be used in tumour immunodiagnosis. Refer to patent for more  
CC sequence analysis details.  
SQ Sequence 18177 BP: 5765 A: 2997 C: 3172 G: 6202 T: 41 Others.

Query Match 11.9%: Score 29.6; DB 1: Length 18177;  
Best Local Similarity 53.4%: Pred. No. 8.8;  
Matches 62: Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 63 GATATGACCATGAACACCTCACAAGTATATCATGAAATGACAAATTTCTTGAT 122  
DB 11042 GACTTATTTAAATGCAAGCTGGGAAAGATATATTAATAGAAAACGCTTGTCATCATCAT 10983  
OY 123 CTCAGACCAACTGATGATCTCTCAAGTAATCTACTGCTCATCCCAA 178  
DB 10982 ATCTAATCTATTTAAATAAATAGCTTTTCCCTGGGAGTATTAAGCTTATTAAGAAA 10927

RESULT 13  
v74512/c

ID V74512 standard; DNA: 6313 BP.  
AC V74512:  
DT 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #201.  
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
OS Staphylococcus aureus.  
FH Location/Qualifiers  
FT key  
FT misc\_feature  
FT 1741..1800  
FT /tag- a  
FT /note- "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
FT misc\_feature  
FT 3541..3600  
FT /tag- b  
FT /note- "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
FT misc\_feature  
FT 5341..5400  
FT /tag- c  
FT /note- "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"  
EP-786519-A2.  
PD 30-JUL-1997.  
PF 07-JAN-1997: 100117.  
PR 05-JAN-1996: US-009861.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
PI Rosen CA;  
DR WPI: 97-374922/35.  
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
PS Claim 1: Page 913-917; 3271pp; English.  
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S.aureus DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against S.aureus infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.  
SQ Sequence 6313 BP: 2226 A: 798 C: 1171 G: 1934 T:

Query Match 11.9%: Score 29.6; DB 1: Length 6313;  
Best Local Similarity 61.8%: Pred. No. 6.1;  
Matches 47: Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 137 CAATGAATCTCTTCAATGATATCTGCTCATCCCAAGAACCAACTGTGAGA 196  
DB 5775 CAAGTACCTTCTAAATCAATATCAACTGCTTATTTGATTCAAATCATCGCTTATGC 5716  
OY 197 AGCTTTTGTGTTAAA 212  
DB 5715 CCTCTTATGTTTATA 5700

```
RESULT 14
X20248_06
Continuation (7 of 10) of X20248 from base 600001 (Borrelia burgdorferi polynucleotide s
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

Query Match 11.9%; Score 29.6; DB 1; Length 110000;
Best Local Similarity 45.9%; Pred. No. 16;
Matches 101; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

OY 28 GTCCTATATCTGACTTGACAGCTCTGGGATGATTATGACATGACAGCTCACA 87
DB 69703 GGCTTAAAGTGAAAATTTTTCATTTCAATCAATCAAAATTTATTCATTCATCCCAAG 69762
OY 88 AGTATATCATTCGATTAAGTACAGATATTCCTGATCTCAGAGCAAGTCAATGAATCTC 147
DB 69763 TGGCATATATATGTAATGCTTCAAAAGACATTAATATACAGATGTTTGTGATTAATCTT 69822
OY 148 TTCAGTGAATACACTCTGCTCATCCCAAGAGCCAACTGAGAGAGTCTTTTGT 207
DB 69823 CTATATATGAAAAAATATGATACAGATCAATTTGTAGGTTAAATGCCCTAATGAGATTGT 69882
OY 208 TTAACCCAGAAAAACATTTCTTTGAAAATGCGACAGATCT 247
DB 69883 CAAAAAATAAAAAATTAGTTTATTAATGCAACAAGTGT 69922

RESULT 15
X20250
X20250 standard; DNA; 111309 BP.
AC X20250:
DE 04-MAY-1999 (first entry)
KW Borrelia burgdorferi polynucleotide sequence #3.
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
FN W0958943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR:
DR WPI: 98-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 738-800; 1128bp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (BD). Products derived from BD can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
```

```
CC Lyme disease.
SQ Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T;

Query Match 11.9%; Score 29.6; DB 1; Length 111309;
Best Local Similarity 45.9%; Pred. No. 16;
Matches 101; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

OY 28 GTCCTATATCTGACTTGACAGCTCTGGGATGATTATGACATGACAGCTCACA 87
DB 80317 GGCTTAAAGTGAAAATTTTTCATTTCAATTCAAATGAATAATATATACATTCATCCCAAG 80376
OY 88 AGTATATCATTCGATTAAGTACAGATATTCCTGATCTCAGAGCAAGTCAATGAATCTC 147
DB 80377 TGGCATATATATGTAATGCTTCAAAAGACATTAATATACAGATGTTTGTGATTAATCTT 80436
OY 148 TTCAGTGAATACACTCTGCTCATCCCAAGAGCCAACTGAGAGAGTCTTTTGT 207
DB 80437 CTATATATGAAAAAATATGATACAGATCAATTTGTAGGTTAAATGCCCTAATGAGATTGT 80496
OY 208 TTAACCCAGAAAAACATTTCTTTGAAAATGCGACAGATCT 247
DB 80497 CAAAAAATAAAAAATTAGTTTATTAATGCAACAAGTGT 80536

Search completed: August 6, 1999, 04:04:15
Job time: 1607 sec
```

Mon Aug 9 13:26:02 1999

us-09-049-696-14.rng

---

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:39:56 ; Search time 3268.17 Seconds

(without alignments)  
228.682 Million cell updates/sec

Title: US-09-049-696-12

Perfect score: 235  
Sequence: 1 GACACGACCAATTCCCCAG.....CAGTGTAAAGTGGGGCTC 235

Scoring table: OLIGO\_NWC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pac:\*  
6: gb\_ph:\*  
7: gb\_p11:\*  
8: gb\_p12:\*  
9: gb\_p13:\*  
10: gb\_p14:\*  
11: gb\_p15:\*  
12: gb\_p16:\*  
13: gb\_p17:\*  
14: gb\_p18:\*  
15: gb\_p19:\*  
16: gb\_p20:\*  
17: gb\_p21:\*  
18: gb\_p22:\*  
19: gb\_p23:\*  
20: gb\_p24:\*  
21: gb\_p25:\*  
22: gb\_p26:\*  
23: gb\_p27:\*  
24: gb\_p28:\*  
25: gb\_p29:\*  
26: gb\_p30:\*  
27: gb\_p31:\*  
28: gb\_p32:\*  
29: gb\_p33:\*  
30: gb\_p34:\*  
31: gb\_p35:\*  
32: gb\_p36:\*  
33: gb\_p37:\*  
34: gb\_p38:\*  
35: gb\_p39:\*  
36: gb\_p40:\*  
37: gb\_p41:\*  
38: gb\_p42:\*  
39: gb\_p43:\*  
40: gb\_p44:\*  
41: gb\_p45:\*  
42: gb\_p46:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	235	100.0	3311	11	AF039400	Homo sapi

2	235	100.0	2826	42	AF127036	AF127036 Homo sapi
3	147	62.6	35278	11	AF039401	AF039401 Homo sapi
4	65	27.7	878	5	195746	195746 Sequence 8
5	20	8.5	216750	1	BSDB0007	299110 Bacillus su
6	20	8.5	130020	42	HUAC004525	AC004525 Homo sapi
7	19	8.1	91201	10	HSDA6819	AT049662 Human DNA
8	18	7.7	76994	7	AB026861	AB026861 Arabidops
9	18	7.7	3604	9	AB026833	AB026833 Homo sapi
10	18	7.7	143875	11	AC002428	AC002428 Human BAC
11	18	7.7	68694	11	AC003676	AC003676 Homo sapi
12	18	7.7	87050	11	AC004763	AC004763 Homo sapi
13	18	7.7	193188	34	AC006884	AC006884 Caenorhab
14	18	7.7	129887	35	AC007064	AC007064 Homo sapi
15	18	7.7	252026	35	AC007174	AC007174 Homo sapi
16	18	7.7	39103	36	CELC3368	U53154 Caenorhabd1
17	18	7.7	150641	36	CEY41E3	295559 Caenorhabd1
18	18	7.7	268369	37	AC005847	AC005847 Drosophila
19	18	7.7	200149	42	AC006548	AC006548 , complet
20	18	7.7	135021	42	AC007159	AC007159 Homo sapi
21	18	7.7	47440	42	U73509	U73509 Homo sapien
22	17	7.2	33818	1	MTCY78	277165 Mycobacteri
23	17	7.2	10010	2	U32709	U32709 Haemophilus
24	17	7.2	2404	5	176892	176892 Sequence 3
25	17	7.2	59762	7	AB023032	AB023032 Arabidops
26	17	7.2	121160	8	ATAC005970	AC005970 Arabidops
27	17	7.2	5940	8	THA133651	AJ133651 Trichoder
28	17	7.2	4964	9	AB002304	AB002304 Human mRN
29	17	7.2	126147	9	HS433M19	295703 Human DNA s
30	17	7.2	5000	9	HUXBCE	L02935 Human major
31	17	7.2	133072	10	HS1097P24	AL035258 Human DNA
32	17	7.2	91201	10	HSDA6819	AL049562 Human DNA
33	17	7.2	152141	10	HSU07000	U07000 Human break
34	17	7.2	3461	10	HSU41060	U41060 Human breas
35	17	7.2	162346	11	AC004551	AC004551 Homo sapi
36	17	7.2	111682	11	AC005162	AC005162 Homo sapi
37	17	7.2	111084	11	AC006486	AC006486 Homo sapi
38	17	7.2	41122	11	HSAC000109	AC000109 Human Cos
39	17	7.2	2937	12	AB011156	AB011156 Mus muscu
40	17	7.2	1755	12	MMCOL18A01	U34606 Mus muscu
41	17	7.2	4514	17	PVB19X559	270593 Parvovirus
42	17	7.2	472	17	PVB19X603	270603 Parvovirus
43	17	7.2	162575	34	AC004086	AC004086 Homo sapi
44	17	7.2	186820	34	AC006400	AC006400 Homo sapi
45	17	7.2	37242	34	AC006673	AC006673 Caenorhab

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS	AF039400					
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hCICL1) mRNA, complete cds.					
ACCESSION	AF039400					
NID	94009457					
VERSION	AF039400.1	GI:4009457				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
AUTHORS	Gruber/A.D., Elble/R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.					
TITLE	Genomic cloning, molecular characterization, and functional analysis of human CICL1, the first human member of the family of Ca2+-activated Cl- channel proteins					
JOURNAL	Genomics 54 (2), 200-214 (1998)					
MEDLINE	99047526					
REFERENCE	2 (bases 1 to 3311)					
AUTHORS	Gruber A.D., Elble, R. and Pauli, B.U.					
TITLE	Direct Submission					

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers  
source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPKSSVFILILHLGALSNSLIQLNNNGYEGIVAVDPNP  
EDFLIQIKDMVTQASLYFEATGKRFYKNAVALIPETWKADIVRPKLEYKNA  
DLVAESTPGNDPEYTEQNGNCEGGERIHLTPDLAKRLAEGPQGAFAHEMAH  
LRMGVDEYNDKEFYLSNGRIQAVRCAGITGNVKKCGGSCYTRKCTFNVTGL  
YKGECEFYLOSROTEKASIMFAQHYDSIVEFCEQNHKREAPKONKONLRTWYI  
RDSDFKRTPTPTOPNPFSLQIGORIVCLVLDKSGMATGNRLRLNQGLFL  
LQTVELGSMVGMATFDSAAVOSSELIQINSQSDRLAKRLPAASGGSISGLLSRA  
FTVIRKKYPTDSEIVLLTDEGNTISGFENEKQSAIITVALGSAOBELELSK  
MTGSLQIVASQVONNGILDAFGALSSNGAVSORSIOLESKGLILQNSQMNVTIV  
DSYVGRDLFLITWTPQIILMDPSQKGGFVVDKNTKMAVLIQIPGIAKGTWY  
SLAASQTLTLTVTSRASNATLPPIVTSKNTKDSKEPPLVYVYANRQKASPLRA  
SVTALIESVNGKTVLELDNGAGADATKDGVSRYFTYTDINGRSYKVALGYN  
AARRVITPOOSGALYIPGWIENDEIOMNPREINKDOVHKOVCSRTSSGSEVAS  
DVPNAPIDLPFGQITLDLKAETHGSLINLTWAPGDYDHGTAKYIIRISTSLD  
LRKFNESLQVNTTALIPKEANSEVLFKRPENITFENGTDLFIQAQVADYDKSEI  
SNIRAVSLFIPQTPPEPSPDTSAPCPNHNISITPGLHILKIMKMWIGELQSLIA  
"

gene  
CDS

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 235; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 1.1e-121;  
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACACGACCAATTCGCCAGCCCTCTGGTAGTTATGCAAAATTCGCCAAGGAGCCTCC 60  
|||  
Db 2149 GACACGACCAATTCGCCAGCCCTCTGGTAGTTATGCAAAATTCGCCAAGGAGCCTCC 2208  
QY 61 CCAATTCAGGGCCAGTGTACAGACCCCTGATTTGAATCAGTGAATGAAAAACAGTATACC 120  
|||  
Db 2209 CCAATTCAGGGCCAGTGTACAGACCCCTGATTTGAATCAGTGAATGAAAAACAGTATACC 2268  
QY 121 TTGGAACACTGAGTAATGAGCAGAGTGTCTGATGCTACTAAGAGAGAGGTGTCTACTCA 180  
|||  
Db 2269 TTGGAACACTGAGTAATGAGCAGAGTGTCTGATGCTACTAAGAGAGAGGTGTCTACTCA 2328  
QY 181 AGGATATTCACAACTTATGACACGAGTGTAGATACAGTGTAAAAAGTGGGGCTC 235  
|||  
Db 2329 AGGATATTCACAACTTATGACACGAGTGTAGATACAGTGTAAAAAGTGGGGCTC 2383  
RESULT 2  
AF127036 2826 bp mRNA PRI 16-Apr-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (Caccl1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
AUTHORS  
TITLE  
1 (bases 1 to 2826)  
Agnel,M., Vermont,T. and Culouscou,J.-M.  
Cloning of three human homologs of bovine epithelial chloride  
channel

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrieres, Rueil-Malmaison 92500, France  
location/Qualifiers  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="Caccl1"  
5. .2749  
/note="Caccl1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="FAD25487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPKSSVFILILHLGALSNSLIQLNNNGYEGIVAVDPNP  
EDFLIQIKDMVTQASLYFEATGKRFYKNAVALIPETWKADIVRPKLEYKNA  
DLVAESTPGNDPEYTEQNGNCEGGERIHLTPDLAKRLAEGPQGAFAHEMAH  
LRMGVDEYNDKEFYLSNGRIQAVRCAGITGNVKKCGGSCYTRKCTFNVTGL  
YKGECEFYLOSROTEKASIMFAQHYDSIVEFCEQNHKREAPKONKONLRTWYI  
RDSDFKRTPTPTOPNPFSLQIGORIVCLVLDKSGMATGNRLRLNQGLFL  
LQTVELGSMVGMATFDSAAVOSSELIQINSQSDRLAKRLPAASGGSISGLLSRA  
FTVIRKKYPTDSEIVLLTDEGNTISGFENEKQSAIITVALGSAOBELELSK  
MTGSLQIVASQVONNGILDAFGALSSNGAVSORSIOLESKGLILQNSQMNVTIV  
DSYVGRDLFLITWTPQIILMDPSQKGGFVVDKNTKMAVLIQIPGIAKGTWY  
SLAASQTLTLTVTSRASNATLPPIVTSKNTKDSKEPPLVYVYANRQKASPLRA  
SVTALIESVNGKTVLELDNGAGADATKDGVSRYFTYTDINGRSYKVALGYN  
AARRVITPOOSGALYIPGWIENDEIOMNPREINKDOVHKOVCSRTSSGSEVAS  
DVPNAPIDLPFGQITLDLKAETHGSLINLTWAPGDYDHGTAKYIIRISTSLD  
LRKFNESLQVNTTALIPKEANSEVLFKRPENITFENGTDLFIQAQVADYDKSEI  
SNIRAVSLFIPQTPPEPSPDTSAPCPNHNISITPGLHILKIMKMWIGELQSLIA  
"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 235; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 1.1e-121;  
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACACGACCAATTCGCCAGCCCTCTGGTAGTTATGCAAAATTCGCCAAGGAGCCTCC 60  
|||  
Db 1802 GACACGACCAATTCGCCAGCCCTCTGGTAGTTATGCAAAATTCGCCAAGGAGCCTCC 1861  
QY 61 CCAATTCAGGGCCAGTGTACAGACCCCTGATTTGAATCAGTGAATGAAAAACAGTATACC 120  
|||  
Db 1862 CCAATTCAGGGCCAGTGTACAGACCCCTGATTTGAATCAGTGAATGAAAAACAGTATACC 1921  
QY 121 TTGGAACACTGAGTAATGAGCAGAGTGTCTGATGCTACTAAGAGATGAGGTGTCTACTCA 180  
|||  
Db 1922 TTGGAACACTGAGTAATGAGCAGAGTGTCTGATGCTACTAAGAGATGAGGTGTCTACTCA 1981  
QY 181 AGGATATTCACAACTTATGACACGAGTGTAGATACAGTGTAAAAAGTGGGGCTC 235  
|||  
Db 1982 AGGATATTCACAACTTATGACACGAGTGTAGATACAGTGTAAAAAGTGGGGCTC 2036  
RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
DEFINITION complete cds.

ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,R.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source 1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="D1 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24532..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310).33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24532..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310).33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PDB:4009460"  
/db\_xref="GI:4009460"  
/translation="MGPEKSSVFLLILHLGALSNSLIOLNNGYEGIVAYIDNPV  
EDELIOIKDMVTQASLYLEATGKRFRFNVAAILIPETKRTADYRRLPYEKKA  
DLVAESTPENGDEPYEOMGNGCEGKERHILTPDFJAGKRLAETGPGKAFHEMAH  
LRMGVFEYNNDEKFTLSNGRIQAVRSGALITGVAKKCGGSCYTKRCYFNVTGL  
YKCEFEVLOSROTEKASIMEQAQVDSIVERCTEQNKGSMATNRLNRLNOAGOLF  
RDSDFKTPPTPTOPNPPTSLQIGRIYCLVDSGSMATNRLNRLNOAGOLF  
LQYELGSMGMAVFDSAAHYOSLQIINSGSDDTLAKRLPAASGGSISCSLSRKA  
FTVIRKPYPTGSEVILLTDGEDNTISCFEYVDSQGLITVTLGSPRANDELISK  
MTGELQYASDOVONNGLIDAFGLSSGNGVNSRISLEKGLILQSSOMKSTVLY  
DSIVGKDTLFLITWTPQQLLWDPGSGQKGGFVNDKNTMALQILPGIAKVTWLY  
SLQASQTLTLTVSRASNATLPITVYSKTKNDTSKFPSPLVVYANRQASPLIRA  
SVTALISVNGKTVTLLELDNGAGADATKDDGVSRFTTDTNGRYSKVALGAVN  
AARRVLPQGGALYIPGMIENDEIOMNPPEIINKDVOVKOYCFSTSGSGSFVAS  
DVNPAPLPDLFPPOITDNLNHIHGSILINTWTAPGDDYDHGTAHNTIYINISIID

LRPKFNSLOVNTALIPKANSSEVLEFPKENTTEENGTDLIAIOAVDKDKSRI  
SNRKNSLFIPQIPPTPSPDSETSAPCPINHINSTITPGIHLIKIMKWIGELQSLTA  
" 6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 62.6% Score 147; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 2 6e-72;  
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GACACGACAAATTCGCCAGCCCTGTGTAATTATGCAAAATATTCGCCAGAGACCTCC 60  
DB 27564 GACACGACAAATTCGCCAGCCCTGTGTAATTATGCAAAATATTCGCCAGAGACCTCC 27623  
QY 61 CCAATTCAGGGCCAGTGTACAGCCCTGATGAAATAGAGCAATGGAAGAAAGCTTACC 120  
DB 27624 CCAATTCAGGGCCAGTGTACAGCCCTGATGAAATAGAGCAATGGAAGAAAGCTTACC 27683  
QY 121 TTGGAATCTGATGATATGAGACAGGT 147  
DB 27684 TTGGAATCTGATGATATGAGACAGGT 27710  
RESULT 4  
LOCUS 195746 878 bp DNA 17-JUL-1998  
DEFINITION Sequence 8 from patent US 5733748.  
ACCESSION 195746  
NID 93940216  
VERSION 195746.1 GI:3940216  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

Unclassified.  
REFERENCE 1 (bases 1 to 878)  
AUTHORS Yu, G. and Rosen, C.  
TITLE Colon specific genes and proteins  
JOURNAL Patent: US 5733748-A 8 31-MAR-1998;  
FEATURES Location/Qualifiers  
source 1..878  
BASE COUNT 257 a 179 c 188 g 241 t 13 others  
ORIGIN /organism="unknown"

Query Match 27.7%; Score 65; DB 5; Length 878;  
Best Local Similarity 100.0%; Pred. No. 3,9e-26;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 TGTCTACTCAGAGTATTCACACTTATGACACGATGATGATACAGTAAAGTCG 230  
Db 1 TGTCTACTCAGAGTATTCACACTTATGACACGATGATGATACAGTAAAGTCG 60  
QY 231 GCCTC 235  
Db 61 GCCTC 65

RESULT 5  
BSUB0007/c  
LOCUS  
DEFINITION Bacillus subtilis complete genome (section 7 of 21): from 1194391 to 1411140.  
ACCESSION 299110 AL009126  
NID 92633472  
VERSION 299110.1 GI:2633472  
KEYWORDS  
SOURCE  
ORGANISM  
Bacillus subtilis.  
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; Bacillus.

1 (bases 1 to 216750)  
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Aebi, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S., Boriss, R., Boudreau, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Entian, K.D., Errington, J., Fader, C., Ferrati, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Gilm, S.Y., Glaser, P., Goffeau, A., Goldlight, E.J., Grenat, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klier, B., Klier, B., Klein, C., Kobayashi, Y., Koetter, P., Konigstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maul, C., Medigue, C., Medina, N., Mellado, R.P., Mazum, K., Moestl, D., Nakai, S., Nopack, M., Noone, D., O'Reilly, M., Ogawa, K., Ogihara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portelle, D., Portolillo, S., Prescott, A.M., Presecan, E., Fujic, P., Purnelle, B., Rapoport, G., Ray, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T., Scanlan, E., Schleich, S., Schroeder, R., Scofield, P., Sekiguchi, J., Sekowska, A., Seror, S.J., Serro, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, K., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassero, A., Viar, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and Danchin, A.

TITLE The complete genome sequence of the gram-positive bacterium  
JOURNAL Bacillus subtilis  
MEDLINE Nature 390 (6657), 249-256 (1997)  
96044033.

REFERENCE 2 (bases 1 to 216750)  
AUTHORS Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.  
TITLE Direct Submission  
JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulated de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, danchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

FEATURES  
source Location/Qualifiers  
1..216750  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
116..1156  
/gene="argC"  
116..1156  
/gene="argC"  
/EC\_number="1.2.1.38"  
/function="arginine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="N-acetylglutamate gamma-semialdehyde dehydrogenase"  
/protein\_id="CAB12976.1"  
/db\_xref="PID:e1183139"  
/db\_xref="PID:g2633473"  
/db\_xref="GI:2633473"  
/db\_xref="SWISS-PROT:P23715"  
/translation="MKIGVATGAGTGTGTLVRLSHHPAECCILYSSSGEGNYSEG YPHNGADOLKPIDMTNRIKHEIDIMPLAARPGVSSSELPPLAAGTIVDLSGLR IEPAYEKRYKRTAPRAVIOENAYGAELEMDQLOQAKLILANPGCFPTAVLLGLAP LQAKLIDSEFVIVDAKTGVSAGKASRGHIFSELDNDFKLYKNEHQHPEIQAL NEMQPLGPTFSALHPMTRGIMATYRLCDLADLDHLYSEFQDSYFVVR KQYDQTEKVEGNSFCIDIAVLTDETRNTIVSVINDIMKGAAGQAVONFNLNMGNE ETLGITITPIYP"  
1176..2396  
/gene="argJ"  
1176..2396  
/gene="argJ"  
/EC\_number="2.3.1.35"  
/EC\_number="2.3.1.1"  
/function="arginine biosynthesis"  
/note="alternate gene name: argH, argE, ornithine acetyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="amino-acid acetyltransferase"  
/protein\_id="CAB12977.1"  
/db\_xref="PID:e1183140"  
/db\_xref="PID:g2633474"  
/db\_xref="GI:2633474"  
/db\_xref="SWISS-PROT:P36843"  
/translation="MIQSEDOIVATGDSVSPKGFQAKGVHCGILRYSKDKLVITSE TPAAVAYVQSHFOAPIKVYODSLKGPILKAVIVSAIANACTGEGQKLDVTRM SEFAVQLEPPLVSSGTIVGELIDMEKIHAGIELKTPAGSGDEPAILITDVI KFOQYEAIGKTYTIGARKSGSMHPRMATHNGFYTTHAIEKRLQALRLETID VSFNKTIVDEGISTDMVLANACAEKLEDEHPDVEFRKALLITCEDLAELIAR DGEGRKLEIAVOGAKNNLDANVIAKIVSNLTKTVYGTDAWNGHIIAIGISAA OVTAEVIVYLGQCLFKNNEPQPSSEIAKYELEGDEIITIVIKMAEDGNGRAMGCD LTVYDKIVNASRT"  
2411..3187  
/gene="argB"  
2411..3187  
/gene="argB"  
/EC\_number="2.7.2.8"  
/function="arginine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="N-acetylglutamate 5-phosphotransferase"  
/protein\_id="CAB12978.1"  
/db\_xref="PID:e1183141"  
/db\_xref="PID:g2633475"  
/db\_xref="GI:2633475"



```
/db_xref="SWISS-PROT:P36840"
/translation="MKTIYFKCGGVITHELEEFYONKEIPASGKWLAIYHGCGPE
ITNMLKRLKITEFGSGGKRTTTPVLEAEMVLSSGVNKFVYAEALRKHGLRAAGISGK
DGLLEADYLDPEYGEVEIKRYASVNAEMNGIIPVIAPIASMTSDCKTLNAD
LAASVAGLEADLMFYVDVDMKORLDVLPKEIOMLIKOEVIIGMIPKYN
ALSALSDOYSEVIVNGKGSFAEDQFOGTRIKVAKKANS"
3184. 4341
/gene="argD"
3184. 4341
/gene="argD"
/EC_number="2.6.1.11"
/function="arginine biosynthesis"
/codon_start=1
/translation="N-acetylornithine aminotransferase"
/protein_id="CAB12979.1"
/db_xref="PID:e1183142"
/db_xref="PID:g2633476"
/db_xref="GI:2633476"
/db_xref="SWISS-PROT:P36839"
/translation="MSLEQTYGRMDIDIRKAGTYVEDQNGKTYLDFIOGIAVSNL
HCEAVTEAVKQDLSVHVSNIQFNSIQEQAQKLAGSADLVFFCNSGGOAEGA
YLRARATGKTKITEFIOSEHGRRTYAGMAATGODKRTGFGMLGGEHYLPYND
KALGEGDIAVAMLETVOSGEGVNPASAEITSAVOSFCKEKALILIDISORTG
KCFAYEHGSLSPDITTVAKGLNGFPVCAVITKQKGEAFPTGSHGTTGGMALAA
VNAITQIVFQDPFLOEADAKGAFLEQLEAEIKSPVKQIRKGLMLGIECDGPVADI
IAELQTLGLVLPAGPERDSAAPLVTDEIAEAVSKLKAIAHSAVNO"
4412. 5473
/gene="cara"
4412. 5473
/EC_number="6.3.5.5"
/function="arginine biosynthesis"
/note="alternate gene name: cpa"
/codon_start=1
/translation="carbamoyl-phosphate transferase-arginine
(subunit A)"
/protein_id="CAB12980.1"
/db_xref="PID:e1183143"
/db_xref="PID:g2633477"
/db_xref="GI:2633477"
/db_xref="SWISS-PROT:P36838"
/translation="MEGYLVLEDGTSFSGELDHGNGEAVFTGMGYOVLIDPS
YKGOIIVTYPLIGYINEXDESKRPQKAAVYEAQDHSHEAVYSLKELYQK
NPLTHYDRAVYKIRAMGTGATVTSKGAETALQDPENVAQASQESTGQD
NKHIALIDEGYKSIASSLVKRGCTVYVYQOMEVYVIRKDGIVLSAGPDPRAIO
PLIKIKSILISREPTLIGICGHOLIAFGGNTFLPCHRGANHPVIDRKRYEMT
SONHSYVDOSINEBELIRFHVNDISVEGLAKKLPVMSVGHPEAHPCAASEM
IFDDYLNKVIIPARRIAHA"
5466. 8549
/gene="carb"
5466. 8549
/gene="carb"
/EC_number="6.3.5.5"
/function="arginine biosynthesis"
/note="alternate gene name: cpa"
/codon_start=1
/translation="carbamoyl-phosphate transferase-arginine
(subunit B)"
/protein_id="CAB12981.1"
/db_xref="PID:e1183144"
/db_xref="PID:g2633478"
/db_xref="GI:2633478"
/db_xref="SWISS-PROT:P18185"
/translation="WPKDTSISISILYIGSGPIITQGAEFYSGTGGCAKEEGRY
LIVNSPATIMTDEAFADETIYEPLESLTIAIKKRPDGLLANGGTALNLEVL
EETGVLEHGVKLLGTSVETIQGEDREFSLMELKPVSESIYVENDADLAFHAE
SIGFVILIRPAYTLGGGGIAPDKAEAFSLMQLASIPINOCLYKESIAQFKEEY
EVMDSNNCTICVKNMENIDPYGHTGDSIVVAPQTLDEDDYOMLRASLTIIISLD
VVGCGNIOFALDPFSKOYVIEVNPVRSVASLQATGYTAKMAAKAVGTYTDEL
KNPLTGSTYASFPALDYIVTFPRMPFDFKPNADRKLGTKAKAIGEVNATIRNDEAA
```

```
IOKASASLEKRNIGTHLPBLSGLSDITLMDLAIITPDDRRFFVYMLLSRSVTDIHE
KTIQDFLHTDITIKLENRLMEAGSDLSLELKRKAKGSDTTLISLSTKEEBY
RLRLKREMOITPSEFKIVDTCAEFDKRTYFSTYEGAEIDGDISREKRALIGSGPI
RIGQGVNEVSAVHGVLTLOELGEPRTIANNPEVSDYEDLADLYEPMTEHIVN
VREGQNEIDFAIVQFGGTAIINAALAEAGITLQTSOTLDVLEDRQFYOLDELG
LKHARGELIATVEEASAKSEIGYGVLLRPSPVIGGMIIYDSOALISLNDSDMS
PYPLIIDGVSKSEYEDLISDGEVFIPYTEHIERAGVSHSGDSFALRPGSITSL
DQGMKDAQKRLARKLTFEYIMTQFVINDGNLTLEVPASRATVPVSKVGVPRAT
```

Query Match 8.5%; Score 20; DB 1; Length 216750;

Best local similarity 100.0%; Pred. No. 0.54;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 GCAAATATTCGCCAGAGAC 56

Db 37659 GCAAATATTCGCCAGAGAC 37640

RESULT 6

HUAC004525

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

On Jun 13, 1998 this sequence version replaced g1:3212842.  
Address all correspondence to: Mark Adams The Institute for Genomic  
Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail  
address: mdadams@tigr.org. The orientation of the sequence is  
from Sp6 end to 3' end. Genes were identified by a combination of  
five methods including: XGRAIL (available by anonymous ftp from  
arthur.epm.ornl.gov), GeneFinder (Phil Green, University of  
Washington), GenScan (Chris Burge,  
http://genomic.stanford.edu/~chris/GENSCANW.html) searches of the  
complete sequence against a peptide database, and the Human gene  
index database at TIGR (http://www.tigr.org/tgd/hgi.html).  
Genes without peptide homology having spliced EST hits are termed  
'Unknown gene product'. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).  
Location/Qualifiers  
1..130020  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16p12"  
/clone="58202"

mRNA complement(Join(74382..76907,78344..78393,80162..80240,  
 83248..83347,86459..86615,92536..92687,94134..94285),  
 96895..97033,98881..98968,108956..109088,125141..125164))  
 STS /note="38310, WI-30149, Chr. 16, Homo sapiens"  
 /db\_xref="dbSTS:G24473"  
 STS 74965..75065  
 /note="12581, WI-9023, Chr. 16, Homo sapiens"  
 /db\_xref="dbSTS:G07084"  
 CDS complement(Join(75329..76907,78344..78393,80162..80240,  
 83248..83347,86459..86615,92536..92687,94134..94285),  
 96895..97033,98881..98950))  
 /codon\_start=1  
 /product="Interleukin 4 alpha-chain precursor"  
 /protein\_id="AAC23495.1"  
 /db\_xref="PID:93219334"  
 /db\_xref="GI:3219334"  
 /translation="MGVLCGLFVSCVLIIVVASSGNMAYLOEPTCVSYMSISTC  
 EMKNGPTNCTELLLIOLVFLSEATCTDENNGAGCYCHLMDVVSADNTLTD  
 LMAGQILMKSEKSEHVPAPGNLTVHNVSDTLITWNPMPDNYLYNLTVA  
 VIMSENPADPRINVTYLEPSRLIASTLSGISYRVRAMACNTWSESPS  
 TKHNSYEPFEOHLLGVSCIVILAVCLCYISIRIKKEMDDQIPNARSILVA  
 IIDAOGSOMKRSRGOEPACPKMNCIFLPCLEHNMKREDDEHRAKEMPEO  
 GSGKAMCPVEISKTIVPESTISVRCVCFEAPVCEEEVEEKEKSRASPESSR  
 DDFORRGIVARLIESFLDLIGENGFCQDMGESCLPPSGSISAHMPDEFPS  
 AGPEAPPWGKEQPLHLEPSPASPTOSDNLCTETPLVLAGNPAYRSFNSLSQSP  
 CPRELGPDLARHLEVEPEKPCVQSEPTVQPEPEWEDILRNVLVHGAASAA  
 PVSAPTSGYOEIVHAEQGTQASAVGLGPEAGYKAFSLASSAVSEKCGFGA  
 SSGEGRYEPFDLIGCPDPAPVPVFTFGLEDREPRSSHLSPSEHGLEP  
 GEKVEDMKPPLPOEATDPLVDSIGSVISALICHLGHLKCHGCHGDCGCPVNA  
 SPCCGCCGDSRSPPTPLRAPSPRGVPLEASLCPSASLAPSGISEKSSSFHRA  
 PGNMOSSQTRKIVNFVSVCPTMYRVS"  
 misc\_feature  
 58000..102000  
 /note="Region: possible breakpoint region"  
 BASE COUNT 35132 a 30950 c 30469 g 33469 t  
 ORIGIN  
 Query Match 8.5%; Score 20; DB 42; Length 130020;  
 Best Local Similarity 100.0%; Pred. No. 0.56;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 137 ATGAGCAGGTGCTGATGCT 156  
 ||||||||||||||||  
 Db 70815 ATGAGCAGGTGCTGATGCT 70834  
 RESULT 7  
 HSDA86H19 91201 bp DNA PRI 10-JUN-1999  
 LOCUS Human DNA sequence from clone 86H19 on chromosome Xq25-26.2,  
 DEFINITION complete sequence.  
 ACCESSION AL049562  
 NID 94958821  
 VERSION AL049562.7 GI:4958821  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 91201)  
 AUTHORS Howden, P.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 On Jun 1, 1999 this sequence version replaced gi:4741573.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

## FEATURES

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Emi, EMBL; Swi, SWISSPROT; Tr, TREMBL  
 IMPORTANT: This sequence is not the entire insert of clone 86H19.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we arrange for a small overlap between  
 neighbouring submissions.  
 The true right end of clone dJ428A13 is at 100 in this sequence.  
 This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.  
 This sequence was generated from part of bacterial clone contigs of  
 human chromosome X, constructed by the Sanger Centre Chromosome X  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/ChrX  
 86H19 is from the library RPi6 constructed at the Roswell Park  
 Cancer Institute by the group of Pieter de Jong. For further  
 details see http://bacpac.med.buffalo.edu/VECTOR: pPi6.  
 location/Qualifiers  
 1. 91201  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /clone="86H19"  
 /map="q25-26.2"  
 /clone\_1bp="RPI6"  
 1. 439  
 /note="L1 repeat: matches 3131..3579 of consensus"  
 420..2037  
 /note="L1P2 repeat: matches 4525..6144 of consensus"  
 2037..2822  
 /note="L1P2 repeat: matches 3730..4515 of consensus"  
 3048..3401  
 /note="L2 repeat: matches 2323..2745 of consensus"  
 3402..3464  
 /note="L1MC4 repeat: matches 7162..7218 of consensus"  
 3653..7355  
 /note="L1M1 repeat: matches -637..3164 of consensus"  
 7356..7426  
 /note="MSRB repeat: matches 356..426 of consensus"  
 7427..7735  
 /note="AluSg repeat: matches 1..302 of consensus"  
 7736..8090  
 /note="MSRB repeat: matches 1..356 of consensus"  
 8091..8600  
 /note="L1M1 repeat: matches -1141..-637 of consensus"  
 8601..9622  
 /note="MER1A repeat: matches 1..1266 of consensus"  
 9623..9863  
 /note="L1M1 repeat: matches -1390..-1141 of consensus"  
 10226..10382  
 /note="L1M1 repeat: matches 5043..5200 of consensus"  
 10390..10523  
 /note="AluCo/FRAM repeat: matches 177..304 of consensus"  
 10525..11116  
 /note="L1M1 repeat: matches 5204..5799 of consensus"  
 11117..11410  
 /note="AluSx repeat: matches 1..294 of consensus"  
 11411..11719  
 /note="L1M1 repeat: matches 5798..6106 of consensus"  
 11733..12905  
 /note="L1M2 repeat: matches -635..677 of consensus"  
 12905..15974  
 /note="L1 repeat: matches 2339..5403 of consensus"  
 17265..17387  
 /note="L2 repeat: matches 2620..2749 of consensus"  
 17456..17570  
 /note="L2 repeat: matches 2582..2709 of consensus"  
 17723..17792  
 /note="35 copies 2 mer ag 766 conserved"

```

repeat_region 17849..18375
/note="MUTCB repeat: matches 3..446 of consensus"
repeat_region 18284..18413
/note="MIR repeat: matches 116..259 of consensus"
repeat_region 18532..18638
/note="L2 repeat: matches 2633..2746 of consensus"
repeat_region 18694..19128
/note="Tiger2a repeat: matches 1..454 of consensus"
repeat_region 19183..19295
/note="L2 repeat: matches 2575..2702 of consensus"
repeat_region 21091..21120
/note="15 copies 2 mer tg 90% conserved"
repeat_region 21122..21296
/note="MIR repeat: matches 57..252 of consensus"
repeat_region 21855..22055
/note="MIR repeat: matches 6..259 of consensus"
repeat_region 22056..22141
/note="43 copies 2 mer ta 73% conserved"
repeat_region 22256..22470
/note="L1P16 repeat: matches 5941..6157 of consensus"
repeat_region 22793..23161
/note="MUT1A2 repeat: matches 2..374 of consensus"
repeat_region 23645..23706
/note="L2 repeat: matches 2623..2683 of consensus"
repeat_region 24163..24328
/note="ALUd repeat: matches 127..287 of consensus"
repeat_region 25148..25353
/note="ALUd repeat: matches 127..287 of consensus"
repeat_region 25673..25982
/note="MIR repeat: matches 47..262 of consensus"
repeat_region 25988..26126
/note="ALU repeat: matches 1..310 of consensus"
repeat_region 26738..27031
/note="L2 repeat: matches 2601..2745 of consensus"
repeat_region 27032..29617
/note="MUT2D repeat: matches 1..291 of consensus"
repeat_region 29618..30184
/note="HERV repeat: matches 3069..5760 of consensus"
repeat_region 30704..31005
/note="ALUd repeat: matches 1..293 of consensus"
repeat_region 31016..31093
/note="ALUd repeat: matches 1..293 of consensus"
repeat_region 32816..32937
/note="39 copies 2 mer ta 65% conserved"
repeat_region 33222..33460
/note="L1P4 repeat: matches 6024..6146 of consensus"
repeat_region 33634..33683
/note="L2 repeat: matches 1539..1795 of consensus"
repeat_region 33937..35713
/note="25 copies 2 mer tt 76% conserved"
repeat_region 35741..36179
/note="L1M8 repeat: matches 4226..6173 of consensus"
repeat_region 36186..36795
/note="L1 repeat: matches 2945..3410 of consensus"
repeat_region 36796..37100
/note="L1P11 repeat: matches -745..-144 of consensus"
repeat_region 37101..42579
/note="ALUsp repeat: matches 1..307 of consensus"
repeat_region 42579..42968
/note="L1P11 repeat: matches -144..6164 of consensus"
repeat_region 42969..43266
/note="L1M1 repeat: matches 2453..2521 of consensus"
repeat_region 43267..43898
/note="ALUsp repeat: matches 1..300 of consensus"
repeat_region 43901..44986
/note="L1M1 repeat: matches 2521..3131 of consensus"
repeat_region 45012..45537
/note="L1P16 repeat: matches 5056..6143 of consensus"
repeat_region 45608..45698
/note="L1 repeat: matches 3125..3690 of consensus"
repeat_region 45724..46459
/note="L1M1 repeat: matches 6238..6325 of consensus"
repeat_region 47298..47415
/note="L1M8 repeat: matches 272..1015 of consensus"

```

```

repeat_region 47829..47898
/note="L1M1 repeat: matches 1..119 of consensus"
repeat_region 48678..48741
/note="35 copies 2 mer at 66% conserved"
repeat_region 49180..49315
/note="ALU/FRAM repeat: matches 226..288 of consensus"
repeat_region 49332..51163
/note="MIR repeat: matches 57..187 of consensus"
repeat_region 53477..53522
/note="L1M4A repeat: matches 4451..6297 of consensus"
repeat_region 54311..54678
/note="23 copies 2 mer ct 76% conserved"
repeat_region 55413..56328
/note="MUTB repeat: matches 10..388 of consensus"
repeat_region 56329..56753
/note="L1M2 repeat: matches 5399..6307 of consensus"
repeat_region 56754..60873
/note="MSTB repeat: matches 1..425 of consensus"
repeat_region 60875..62641
/note="L1M2 repeat: matches 1255..5399 of consensus"
repeat_region 62917..63383
/note="L1M1 repeat: matches -981..757 of consensus"
repeat_region 63404..63796
/note="MER65A repeat: matches 1..445 of consensus"
repeat_region 63799..65091
/note="MSTB repeat: matches 1..424 of consensus"
repeat_region 65115..65298
/note="MSTB-internal repeat: matches 1..1320 of consensus"
repeat_region 65115..65298
/note="MSTB-internal repeat: matches 1470..1651 of consensus"

Query Match 8.1%; Score 19; DB 10; Length 91201;
Best Local Similarity 100.0%; Pred.No. 2.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 TTATGACGATGGTAGA 213
DB 81423 TTATGACGATGGTAGA 81441

RESULT 8
AB026661/c 76994 bp DNA PLN 07-MAY-1999
LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: T3066,
DEFINITION complete sequence.
ACCESSION AB026661
NID 94757417
VERSION AB026661.1 GI:4757417
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui PI
clone:T3066.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
REFERENCE 1 (sites)
AUTHORS Nakamura,Y.
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 5. XIII
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 76994)
AUTHORS Nakamura,Y.
TITLE Direct Submision
JOURNAL Submitted (28-APR-1999) to the DDBJ/EMBL/Genbank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2; Yana 1532-3, Kisarazu, Chiba 292-0812 Japan
(E-mail:yakam@kazusa.or.jp, Tel:81-438-52-3935(ex.7443),
Fax:81-438-52-3934)
FEATURES
1..76994
Location/Qualifiers
source
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"

```

/chromosome="5"  
/clone="T306"  
/clone\_lib="Mitsui P1"  
BASE COUNT 25435 a 12888 c 13110 g 2561 t  
ORIGIN

Query Match 7.7%; Score 18; DB 7; Length 76994;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 CAATATTCGCCAGAG 55  
|||||  
DB 75006 CAATATTCGCCAGAG 74989

## RESULT 9

AB026833 3604 bp mRNA PRI 26-MAY-1999  
LOCUS Homo sapiens mRNA for chloride channel protein, complete cds.  
DEFINITION  
ACCESSION AB026833  
NID 94887600  
VERSION AB026833.1 GI:4887600  
KEYWORDS chloride channel protein.  
SOURCE Homo sapiens adult corneal epithelium cDNA to mRNA,  
clone\_lib:lambda ZAP clone:lambda 13802.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (sites)  
Itoh, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.  
Isolation and characterization of chloride channel in human corneal  
epithelium  
Unpublished (1999)  
2 (bases 1 to 3604)  
Itoh, R., Kawamoto, S. and Okubo, K.  
Direct Submission  
Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases. Reiko  
Itoh, Institute for Molecular and Cellular Biology, Osaka  
University, Yamada-oka 1-3, Suita, Osaka 565-0871, Japan  
(E-mail: reiko@imcb.osaka-u.ac.jp, Tel: 81-6-6879-7992,  
Fax: 81-6-6877-1922)

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

FEATURES  
Source

## CDS

1..3604  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="lambda 13802"  
/clone\_lib="lambda ZAP"  
/dev\_stage="adult"  
/map="jp32"  
/tissue\_type="corneal epithelium"  
42..2873  
/codon\_start=-1  
/product="chloride channel protein"  
/protein\_id="BAA7810.1"  
/db\_xref="pid:d1041578"  
/db\_xref="pid:g4887601"  
/db\_xref="gi:4887601"  
/translation="MTORSTAGPICNIKFTLLVALSSSELPYAGAVQLODNGYNGILL  
IAINPOVVENONLISNIKEMITSEAFYLNATRRVFFRNKILIPATWANNRSKIK  
QESYERKANVITDMYGAGHDDPYTLQRCGKEGKIYHFNFLINDNLNLAGSGSR  
VFVHEMHLRMGVDEYINDKPYINGONQIKVTCSSDITGLFVCEKGCPCENCI  
SKLFKEGCTFVNSTONATASIMFMOSSVFEPCNASTNOCAPNOMOCMLBRAM  
DYITDSADPHHSFPMNTELPPTPTSLVAGKVVCLVDVSSKMAEDRLQLDQA  
AEYTLAQIVIEHTFVGIAFSDSGEIRAOHQINSNDRLVSYLPTVSATDLSI  
GSGKGFVEVERKNGKAGSVMLVTSGDKLGNCLPTVLSGSTIHSIALGSSAA  
PNEELSLRLTGLKFEVPDISNSMIDAFSRISGSGDITFOOHIQLESSEGVKHH  
OLKNTYVDNTVGNDFELVWQASGPPEILFDPGKRYTNTFTTFRASMT  
PGTAKGHTYTLNTHHSIALKVYTSRASNSAVPRATVEVERDSIHPHPYMI  
YAVKOGFYIILNATYATYEPETGDPYTRLLDDAGADVINKDGIISRTFSFAN  
GRISLKVHNVHSPISIPSHSIPGSHAMVGYGTANGNIOQMNPRKSVGNEEERWNG  
FSHVSOGSGSFVIGVAPHPDVPCKIIDEAVVEBELTISMTAPGDEPDGOGAT

SYEIRMSKSLQNTODDENNALIVNTESKRNPOAGIREIFTFSPQISTNGEPHONGET  
HESHRIYVAIRADNRNSLOASVNIQAFLPIPPNSDPVAPARDYLLIKGLVFMGLIG  
ICILIVVTHHTLSRKRKRDRENGTKLL"  
BASE COUNT 1113 a 740 c 732 g 1017 t 2 others  
ORIGIN

Query Match 7.7%; Score 18; DB 9; Length 3604;  
Best Local Similarity 100.0%; Pred. No. 9.1;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 ATGAGCAGGTGCTGATG 154  
|||||  
DB 2017 ATGAGCAGGTGCTGATG 2034

## RESULT 10

AC002428 143875 bp DNA PRI 19-DEC-1997  
LOCUS Human BAC clone GS039E22 from 5q31, complete sequence.  
DEFINITION  
ACCESSION AC002428  
NID 92335068  
VERSION AC002428.1 GI:2335068  
KEYWORDS HMG.  
SOURCE human.  
ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 143875)  
Strong, C., Biewald, T., Jin-Wollam, A and Duckels, G.  
The sequence of H. sapiens BAC clone GS039E22  
Unpublished (1997)  
2 (bases 1 to 143875)  
Waterston, R.

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Submitted (19-NOV-1997) Department of Genetics, Washington  
University 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
Submitted By:  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:saplense@wustl.edu

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone; and the assembly was  
confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University School of  
Medicine, St. Louis MO.

## SOURCE INFORMATION:

This clone is from the first BAC library from Genome Systems, Inc.  
(http://www.genomesystems.com).  
Cell line: lymphoblastoid  
Haplotypes: two  
VECTOR: pBelobAC  
Selection: chloramphenicol

## NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of GS039E22;  
actual end is at 143875 of GS039E22. The orientation of this clone

```
FEATURES             is unknown.
Source                Location/Qualifiers
1..143875
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="GS039E22"
/clone_1lb="GSBAC1"
/map="5q31"
complement(5925..6381)
repeat_region
/rpt_family="L1"
6395..7261
/rpt_family="L1"
complement(9580..10034)
repeat_region
/rpt_family="ALU"
15719..16011
/rpt_family="ALU"
16686..16718
/rpt_family="L1"
17851..17899
/rpt_family="L1"
22012..22036
/rpt_family="L1"
22113..22221
/rpt_family="L1"
24193..24236
/rpt_family="L1"
30275..31887
/rpt_family="L1"
33104..33327
/rpt_family="ALU"
33244..33528
/rpt_family="ALU"
34799..37445
/rpt_family="L1"
complement(36012..36430)
repeat_region
/rpt_family="L1"
38337..38815
/rpt_family="ALU"
38955..39081
/rpt_family="L1"
complement(40756..41047)
repeat_region
/rpt_family="ALU"
complement(41119..41151)
/rpt_family="L1"
48982..49122
/rpt_family="L1"
49941..50794
/rpt_family="L1"
50979..51144
/rpt_family="L1"
complement(51255..51279)
/rpt_family="L1"
51518..52173
/rpt_family="L1"
52352..53446
/rpt_family="L1"
complement(57647..57935)
/rpt_family="ALU"
complement(58385..59161)
/rpt_family="L1"
complement(59607..59887)
/rpt_family="ALU"
complement(59979..60046)
/rpt_family="L1"
complement(61870..62160)
/rpt_family="ALU"
complement(62446..62494)
/rpt_family="L1"
complement(62565..62857)
/rpt_family="ALU"
64970..65148
/note="match to human EST H53181 (NID:g993328) Yq83f08.r1"

gene
CDS
65103..72605
/gene="WUSC:GS039E22.2"
join(65103..65148,66969..67065,68568..68713,72439..72605)
/note="WUSC:GS039E22.2"
/note="similar to chicken myeloid protein-1 precursor; 60%
similarity to P08940 (PID:g127095)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB6905.1"
/db_xref="PID:g2335070"
/db_xref="GI:2335070"
/translation="MESTKALLAGLISTALAGPMANICAGSSNEITCDRHGGGV
SAORSORHGVGDIAGSVYAPFTGMIVGDEKPYONKNAINNGVIRISGRGCVKM
FYIRPIKXKPKIKGKGLTLPLQKYPPIGSIQSHVHENCSSDPATL"
65477..65504
/rpt_family="L1"
66438..66725
/rpt_family="ALU"
66968..67065
/gene="WUSC:GS039E22.2"
/note="match to human EST H53181 (NID:g993328) Yq83f08.r1"
68366..68694
/gene="WUSC:GS039E22.2"
/note="match to human EST H53181 (NID:g993328) Yq83f08.r1"
complement(68868..69019)
/rpt_family="L1"
69424..69451
/rpt_family="L1"
complement(70815..70874)
/rpt_family="L1"
70962..71252
/rpt_family="ALU"
complement(72301..72364)
/rpt_family="L1"
72683..72819
/rpt_family="L1"
73775..73842
/rpt_family="L1"
complement(76149..76437)
/rpt_family="ALU"
complement(77820..77864)
/rpt_family="L1"
complement(79481..79517)
/rpt_family="L1"
complement(80014..80341)
/note="match to human EST Z45085 (NID:g574280)"
83779..83817
/rpt_family="L1"
85736..85928
/rpt_family="ALU"
85960..86088
/rpt_family="ALU"
86112..86266
/rpt_family="ALU"
86269..86302
/rpt_family="L1"
88547..88584
/rpt_family="L1"
complement(89933..90211)
/rpt_family="ALU"
complement(90265..90566)
/rpt_family="ALU"
complement(91084..91505)
/rpt_family="L1"
complement(91775..92785)
/rpt_family="L1"
complement(93324..93694)
/rpt_family="MER"
complement(95492..95806)
/rpt_family="MER"
complement(95891..96119)
/rpt_family="MER"
```

Query Match 7.7% Score 18; DB 11; Length 143875;

Best Local Similarity 100.0% Freq. No. 7.4;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 ATTGATCAGTGAATGCA 108

Db 142316 ATTGATCAGTGAATGCA 142333

RESULT 11

AC003676

LOCUS AC003676 86894 bp DNA PRI 11-DEC-1997

DEFINITION Homo sapiens chromosome 2, P1 clone 777H5 (BRL H27), complete sequence.

ACCESSION

AC003676 AC002198 L81789 L81625 L49047 L46908 L77047 L77048 L78757  
L81626 L81790 AC000964 AC000963 AC002200 AC002202 AC002201 AC000962  
AC002199 AC001484 AC001483 AC000965 L81787 L81624 L78755 L78756  
L81788 L81785 L81786 L81791 L81627 AC000961 L81784

NTID

AC003676.1 GI:2677682

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

/note="(AC)18"

/rpt\_type=tandem

/rpt\_unit=AC

10006..10576

/rpt\_family="Alu"

10577..10643

/rpt\_family="LOR1"

/note="(LOR1)"

complement(11015..11089)

/note="GRAIL 2 good exon, frame 0"

11691..11985

/rpt\_family="AluSx"

12098..13206

/rpt\_family="MLT1a"

complement(13245..13358)

/note="GRAIL 2 good exon, frame 0"

complement(13774..14065)

/rpt\_family="Alu"

complement(14402..14660)

/rpt\_family="L1"

15262..15285

/note="(GT)12"

/rpt\_type=tandem

/rpt\_unit=GT

15486..15576

/rpt\_family="L1"

15803..16088

/rpt\_family="L1"

16284..17121

/rpt\_family="L1"

17234..17325

/rpt\_family="L1"

18334..19376

/rpt\_family="MER42"

complement(19400..19787)

/rpt\_family="THER1"

19856..20253

/rpt\_family="Alu"

22779..23073

/rpt\_family="Alu"

23074..23310

/rpt\_family="Alu"

24666..25002

/rpt\_family="MER7"

complement(25524..25829)

/rpt\_family="Alu"

complement(26101..26647)

/rpt\_family="L1"

26845..27019

/note="GRAIL 2 good exon, frame 2"

complement(27832..27904)

/rpt\_family="MIR"

29398..29574

/note="GRAIL 2 marginal exon, frame 0"

complement(30287..30515)

/rpt\_family="L1"

30571..30685

/note="GRAIL 2 good exon, frame 2"

complement(31720..31836)

/note="81% identity -- 275892"

complement(32050..32128)

/note="81% identity -- 275892"

complement(32142..33043)

/rpt\_family="L1"

complement(33322..33616)

/note="GRAIL 2 marginal exon, frame 1"

33840..34055

/rpt\_family="MER30"

complement(34512..35358)

/rpt\_family="L1"

complement(35714..35973)

/rpt\_family="L1"

36143..36268

/rpt\_family="MIR2"

```
misc_feature complement(36503..36637)
      /note="GRAIL 2 good exon, frame 1"
repeat_region complement(39724..40756)
      /rpt_family="LI"
misc_feature complement(40758..41018)
      /note="GRAIL 2 good exon, frame 0"
      42234..42339
misc_feature /note="GRAIL 2 marginal exon, frame 0"
      complement(42575..42851)
      /rpt_family="MIR"
repeat_region complement(43332..43437)
      /rpt_family="LI"
repeat_region .44300..45097
      /rpt_family="MER42"
repeat_region 45192..48982
      /rpt_family="LI"
      complement(50017..50090)
repeat_region /rpt_family="Alu"
      complement(50431..50724)
repeat_region /rpt_family="Alu"
      complement(50725..51120)
      /rpt_family="Alu"
repeat_region 51691..52300
      /rpt_family="LI"
repeat_region 52567..53200
      /rpt_family="LI"
repeat_region 53114..53490
      /rpt_family="LI"
      complement(53837..53945)
repeat_region /rpt_family="Alu"
      complement(54168..54649)
repeat_region /rpt_family="Alu"
      complement(54650..54936)
repeat_region /rpt_family="Alu"
      complement(55073..55504)
      /rpt_family="LI"
repeat_region complement(55944..56013)
      /rpt_family="LI"
      56113..56140
      /note="(T)28"
repeat_region /rpt_type=tandem
      /rpt_unit=1
      complement(56114..56434)
repeat_region /rpt_family="LI"
      complement(56955..57143)
repeat_region /rpt_family="LI"
      complement(57846..57924)
      /note="88 identity -- T91321"
      /db_xref="dbEST:T91321"
      complement(58001..58124)
      /note="83% identity -- T91321"
      /db_xref="dbEST:T91321"
      61272..61467
      /rpt_family="Alu"
repeat_region /rpt_family="Alu"
      complement(62102..64322)
      /rpt_family="LI"
repeat_region 65002..65166
      /rpt_family="LI"
      65221..65442
      /rpt_family="LI"
      /rpt_family="LI"
      65454..65686
      /rpt_family="MSTB"
      65657..65680

Query Match 7.7% Score 18; DB 11; Length 86894;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 12
AC004763
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
COMMENT
FEATURES
source
misc_feature
repeat_region
repeat_region
gene
repeat_region
misc_feature
CDS
repeat_region
prim_transcript

AC004763 87050 bp DNA PRI 03-SEP-1998
Hom sapiens chromosome 5, P1 clone 854b11 (LBNU H44), complete
sequence.
AC004763 AC002220 AC001019 AC001020 AC001021 AC001022 AC001023
AC001503 AC001504 AC001505 AC001506 AC001507 LB1403 LB1404 LB1658
LB1659 LB1660 LB1843 LB1844 LB1845 LB1846
93168621
AC004763.1 GI:3168621
HTG.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 87050)
Kimberly, N., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Kadner, K., Miguel, T., Miller, C., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
Sequencing of human chromosome 5
Unpublished
2 (bases 1 to 87050)
Ricke, D.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
Unpublished
3 (bases 1 to 87050)
Kimberly, N., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
Direct Submission
Submitted (29-MAY-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
Sequence submitted by:
DOE Joint Genome Institute.
Location/Qualifiers
1..87050
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5q"
/clone="854b11"
/chromosome="5"
/note="LBNU H44"
1..83788
/note="98% identity AC002428 BAC clone GS039E22"
complement(1692..2035)
/rpt_family="Alu"
complement(2363..2734)
/rpt_family="Alu"
2598..13729
/gene="LEC2"
complement(3888..4290)
/rpt_family="THE1"
3961..4058
/gene="LEC2"
/note="GRAIL 2 excellent exon, frame 1"
join(4949..4994,6815..6911,8414..8559,12285..12451)
/join(4949..4994,6815..6911,8414..8559,12285..12451)
/gene="LEC2"
/codon_start=1
/evidence=experimental
/product="LEC2"
/protein_id="AAC17734.1"
/db_xref="PID:93168622"
/db_xref="GI:3168622"
/translatability="WSTKALLAGLSTALAGPWANICAGKSSNEIRCDHCCGOY
SAQSORHOGCVLCSGSGTVAPFTGMIVGGRPYONKNAINNVRISGRPCVVM
FYIRPKYKGIKRGKRGTLPLQKYVPGIOSHVHIENCDSDPTAVL"
6265..6589
/rpt_family="Alu"
prim_transcript join(8414..8559,12285..12435)
```

```

/gene="IECT2"
/standard_name="mj49e04"
/note="81% identity mouse EST dbEST:AA049146"
repeat_region
/rpt_family="I1"
complement(8716..8871)
/rpt_family="I1"
10772..11105
/rpt_family="Alu"
complement(11123..11370)
/rpt_family="MER33"
complement(11273..11373)
/rpt_family="MER3"
11337..11370
/rpt_family="MER3"
12537..12665
/rpt_family="L1MA2"
complement(115974..16294)
/rpt_family="Alu"
19943..20049
/rpt_family="MIR2"
complement(22248..22370)
/note="GRAIL 2 excellent exon, frame 0"
misc_feature
complement(22835..22991)
/note="GRAIL 2 excellent exon, frame 1"
repeat_region
/rpt_family="MIR"
23708..23822
/rpt_family="MIR"
25596..26149
/rpt_family="Alu"
26527..26687
/note="GRAIL 2 excellent exon, frame 2"
misc_feature
27183..27329
/note="GRAIL 2 excellent exon, frame 0"
complement(29787..30416)
/rpt_family="Alu"
30082..30105
/note="(CTT)6"
/rpt_type="tandem"
/rpt_unit="CTT"
30103..30131
/note="(T)29"
/rpt_type="tandem"
/rpt_unit="T"
30986..31099
/note="GRAIL 2 excellent exon, frame 1"
complement(31614..32740)
/rpt_family="I1"
complement(32960..33553)
/rpt_family="MER39B"
35335..37690
/rpt_family="T1gger1"
39895..40311
/standard_name="UTR819"
/note="98% identity"
/db_xref="dbSTS:LI8431"
complement(40348..40385)
/note="GRAIL 2 excellent exon, frame 0"
42100..42237
/rpt_family="MIR2"
46118..46419
/rpt_family="Alu"
complement(49213..49374)
/note="GRAIL 2 excellent exon, frame 1"
50864..50953
/rpt_family="MIR"
51489..51646
/rpt_family="MLT1"
51980..52321
/rpt_family="MER42"
52476..52726
/rpt_family="MER42"
complement(52724..53048)
/rpt_family="Alu"
53021..53116
/rpt_family="MER42a"

```

```

repeat_region 54201..54264
/rpt_family="MIR"
complement(54269..54348)
/rpt_family="MADE1"
55109..55419
/rpt_family="Alu"
complement(56337..56579)
/rpt_family="L1PB3"
57726..57846
/rpt_family="MLTc"
complement(61064..61158)
/rpt_family="MIR"
62546..62691
/rpt_family="MIR"
62776..62857
/rpt_family="MIR"
62869..63188
/rpt_family="Alu"
63074..67737
/gene="IL-9"
/note="Interleukin-9"
/evidence=experimental
63151..63171
/note="(A)21"
/rpt_type="tandem"
/rpt_unit="A"
join(63988..64089,64202..64237,64320..64352,65641..65772,
67284..67403)
/gene="IL-9"
/note="Interleukin-9"
/codon_start=1
/evidence=experimental
/product="IL-9"
/protein_id="AAC17735.1"
/db_xref="PIR:G3168623"
/db_xref="GI:3168623"
/translacion="MYTTSALLLCSYAGGCGCTIAGIILINFLINKMODPASKCHS
ANTVSCICGIPSDNCTRPCFSERLSQNTIMQRYPLIFSRVKSVEVLNNKCPY
FSCQPCNQITAGNALTLFLKSLLEIFQEKRMGMCKI"
complement(66140..66462)
/rpt_family="Alu"
complement(66682..66902)
/rpt_family="MIR"
67013..67060
/note="(GT)24"
/rpt_type="tandem"
/rpt_unit="GT"
complement(68000..68073)
/rpt_family="MIR"
complement(69079..69155)
/rpt_family="MIR"
69608..69913
/rpt_family="Alu"
complement(70060..70300)
/rpt_family="I1"
71152..71421
/standard_name="WT-13437"
/db_xref="dbSTS:G23760"

```

```

Query Match 7.7% Score 18; DB 11; Length 87050;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 ATTGAATCAGTGAATGGA 108
Db 82212 ATTGAATCAGTGAATGGA 82229

RESULT 13
AC006884 193188 bp DNA HTG 26-FEB-1999
LOCUS Caenorhabditis elegans clone Y57E12, WORKING DRAFT SEQUENCE, 4
DEFINITION unordered pieces.

```



```

ACCESSION      AC006884
NTD            AC0309911
VERSION        AC006884.2  GI:4309911
KEYWORDS       HTG; HTGS_PHASE1.
SOURCE         Caenorhabditis elegans.
ORGANISM       Caenorhabditis elegans.
REFERENCE      Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
AUTHORS        Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis
TITLE          1 (bases 1 to 193188)
JOURNAL        Waterston, R.H.
REFERENCE      The sequence of Caenorhabditis elegans clone
AUTHORS        Unpublished
TITLE          2 (bases 1 to 193188)
JOURNAL        Waterston, R.H.
REFERENCE      Direct Submission
AUTHORS        Submitted (24-FEB-1999) Genome Sequencing Center, Washington
TITLE          University School of Medicine, 4444 Forest Park Parkway, St. Louis
JOURNAL        MO 63108, USA
COMMENT        MO Mar 1, 1999 this sequence version replaced gi:4263464.
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 4 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *
                * 1
                * 2784: contig of 2784 bp in length
                *
                * 2785
                * 2799: gap of unknown length
                *
                * 2800
                * 37804: contig of 35005 bp in length
                *
                * 37805
                * 37819: gap of unknown length
                *
                * 37820
                * 74165: contig of 36346 bp in length
                *
                * 74166
                * 74180: gap of unknown length
                *
                * 74181
                * 193188: contig of 119008 bp in length.
                *
                * Location/Qualifiers
                *
                * 1..193188
                * /organism="Caenorhabditis elegans"
                * /db_xref="taxon:6239"
                * /clone="Y57E12"
BASE COUNT     63610 a 33574 c 33884 g 62075 t 45 others
ORIGIN
Query Match    7.7%; Score 18; DB 34; Length 193188;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

cy 101 TGAATGAAAAACAGTTA 118
Db 142043 TGAATGAAAAACAGTTA 142060

RESULT 14
LOCUS      AC007064/c
DEFINITION Homo sapiens, WORKING DRAFT SEQUENCE, 16 unordered pieces.
ACCESSION  AC007064
NID        95080831
VERSION    AC007064.11  GI:5080831
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homiidae; Homo.
            1 (bases 1 to 129887)
            Ray, L.A.; Loh, P.; Qi, S.; Sloan, D.; McDermid, H. and Roe, B.A.
            Homo sapiens Chromosome 22q11 PAC Clone p8708 in CES Region
            Unpublished
            2 (bases 1 to 129887)
            Ray, L.A.; Loh, P.; Qi, S.; Sloan, D.; McDermid, H. and Roe, B.A.
            Direct Submission
            Submitted (11-MAR-1999) Department of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

```

```

COMMENT
OR 73019, USA
ON Jun 16, 1999 this sequence version replaced gi:50659480.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2019: contig of 2019 bp in length
* 2020 2083: gap of unknown length
* 2084 5505: contig of 3422 bp in length
* 5506 5569: gap of unknown length
* 5570 11532: contig of 5963 bp in length
* 11533 11596: gap of unknown length
* 11597 17981: contig of 6385 bp in length
* 17982 18045: gap of unknown length
* 18046 20114: contig of 2069 bp in length
* 20115 20178: gap of unknown length
* 20179 25453: contig of 5277 bp in length
* 25456 25519: gap of unknown length
* 25520 38800: contig of 13281 bp in length
* 38801 38864: gap of unknown length
* 38865 42442: contig of 3578 bp in length
* 42443 42506: gap of unknown length
* 42507 48011: contig of 5505 bp in length
* 48012 48075: gap of unknown length
* 48076 56063: contig of 7988 bp in length
* 56064 56127: gap of unknown length
* 56128 63600: contig of 7473 bp in length
* 63601 63664: gap of unknown length
* 63665 71730: contig of 8066 bp in length
* 71731 71794: gap of unknown length
* 71795 93933: contig of 22139 bp in length
* 93934 93997: gap of unknown length
* 93998 106689: contig of 12692 bp in length
* 106690 106753: gap of unknown length
* 106754 119337: contig of 12583 bp in length
* 119338 129887: contig of 10487 bp in length.
* 129888 129901: contig of 10487 bp in length.
* 129902 129915: contig of 10487 bp in length.
* 129916 129929: contig of 10487 bp in length.
* 129930 129943: contig of 10487 bp in length.
* 129944 129957: contig of 10487 bp in length.
* 129958 129971: contig of 10487 bp in length.
* 129972 129985: contig of 10487 bp in length.
* 129986 129999: contig of 10487 bp in length.
* 130000 130013: contig of 10487 bp in length.
* 130014 130027: contig of 10487 bp in length.
* 130028 130041: contig of 10487 bp in length.
* 130042 130055: contig of 10487 bp in length.
* 130056 130069: contig of 10487 bp in length.
* 130070 130083: contig of 10487 bp in length.
* 130084 130097: contig of 10487 bp in length.
* 130098 130111: contig of 10487 bp in length.
* 130112 130125: contig of 10487 bp in length.
* 130126 130139: contig of 10487 bp in length.
* 130140 130153: contig of 10487 bp in length.
* 130154 130167: contig of 10487 bp in length.
* 130168 130181: contig of 10487 bp in length.
* 130182 130195: contig of 10487 bp in length.
* 130196 130209: contig of 10487 bp in length.
* 130210 130223: contig of 10487 bp in length.
* 130224 130237: contig of 10487 bp in length.
* 130238 130251: contig of 10487 bp in length.
* 130252 130265: contig of 10487 bp in length.
* 130266 130279: contig of 10487 bp in length.
* 130280 130293: contig of 10487 bp in length.
* 130294 130307: contig of 10487 bp in length.
* 130308 130321: contig of 10487 bp in length.
* 130322 130335: contig of 10487 bp in length.
* 130336 130349: contig of 10487 bp in length.
* 130350 130363: contig of 10487 bp in length.
* 130364 130377: contig of 10487 bp in length.
* 130378 130391: contig of 10487 bp in length.
* 130392 130405: contig of 10487 bp in length.
* 130406 130419: contig of 10487 bp in length.
* 130420 130433: contig of 10487 bp in length.
* 130434 130447: contig of 10487 bp in length.
* 130448 130461: contig of 10487 bp in length.
* 130462 130475: contig of 10487 bp in length.
* 130476 130489: contig of 10487 bp in length.
* 130490 130503: contig of 10487 bp in length.
* 130504 130517: contig of 10487 bp in length.
* 130518 130531: contig of 10487 bp in length.
* 130532 130545: contig of 10487 bp in length.
* 130546 130559: contig of 10487 bp in length.
* 130560 130573: contig of 10487 bp in length.
* 130574 130587: contig of 10487 bp in length.
* 130588 130601: contig of 10487 bp in length.
* 130602 130615: contig of 10487 bp in length.
* 130616 130629: contig of 10487 bp in length.
* 130630 130643: contig of 10487 bp in length.
* 130644 130657: contig of 10487 bp in length.
* 130658 130671: contig of 10487 bp in length.
* 130672 130685: contig of 10487 bp in length.
* 130686 130699: contig of 10487 bp in length.
* 130700 130713: contig of 10487 bp in length.
* 130714 130727: contig of 10487 bp in length.
* 130728 130741: contig of 10487 bp in length.
* 130742 130755: contig of 10487 bp in length.
* 130756 130769: contig of 10487 bp in length.
* 130770 130783: contig of 10487 bp in length.
* 130784 130797: contig of 10487 bp in length.
* 130798 130811: contig of 10487 bp in length.
* 130812 130825: contig of 10487 bp in length.
* 130826 130839: contig of 10487 bp in length.
* 130840 130853: contig of 10487 bp in length.
* 130854 130867: contig of 10487 bp in length.
* 130868 130881: contig of 10487 bp in length.
* 130882 130895: contig of 10487 bp in length.
* 130896 130909: contig of 10487 bp in length.
* 130910 130923: contig of 10487 bp in length.
* 130924 130937: contig of 10487 bp in length.
* 130938 130951: contig of 10487 bp in length.
* 130952 130965: contig of 10487 bp in length.
* 130966 130979: contig of 10487 bp in length.
* 130980 130993: contig of 10487 bp in length.
* 130994 131007: contig of 10487 bp in length.
* 131008 131021: contig of 10487 bp in length.
* 131022 131035: contig of 10487 bp in length.
* 131036 131049: contig of 10487 bp in length.
* 131050 131063: contig of 10487 bp in length.
* 131064 131077: contig of 10487 bp in length.
* 131078 131091: contig of 10487 bp in length.
* 131092 131105: contig of 10487 bp in length.
* 131106 131119: contig of 10487 bp in length.
* 131120 131133: contig of 10487 bp in length.
* 131134 131147: contig of 10487 bp in length.
* 131148 131161: contig of 10487 bp in length.
* 131162 131175: contig of 10487 bp in length.
* 131176 131189: contig of 10487 bp in length.
* 131190 131203: contig of 10487 bp in length.
* 131204 131217: contig of 10487 bp in length.
* 131218 131231: contig of 10487 bp in length.
* 131232 131245: contig of 10487 bp in length.
* 131246 131259: contig of 10487 bp in length.
* 131260 131273: contig of 10487 bp in length.
* 131274 131287: contig of 10487 bp in length.
* 131288 131301: contig of 10487 bp in length.
* 131302 131315: contig of 10487 bp in length.
* 131316 131329: contig of 10487 bp in length.
* 131330 131343: contig of 10487 bp in length.
* 131344 131357: contig of 10487 bp in length.
* 131358 131371: contig of 10487 bp in length.
* 131372 131385: contig of 10487 bp in length.
* 131386 131399: contig of 10487 bp in length.
* 131400 131413: contig of 10487 bp in
```

TITLE Homo sapiens, clone hRPK\_93\_A\_1  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 252026)  
AUTHORS Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,T., Karatas,A., Lehoczy,J., Liu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGuck,A., McKernan,K., McLaughlin,J., Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomas,N., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Toriella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (26-MAR-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

\* NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1. 3174: contig of 3174 bp in length  
\* gap of unknown length  
\* 3175 6910: contig of 3736 bp in length  
\* gap of unknown length  
\* 6911 10684: contig of 3774 bp in length  
\* gap of unknown length  
\* 10685 14638: contig of 3954 bp in length  
\* gap of unknown length  
\* 14639 17903: contig of 3265 bp in length  
\* gap of unknown length  
\* 17904 24493: contig of 6590 bp in length  
\* gap of unknown length  
\* 24494 29359: contig of 4866 bp in length  
\* gap of unknown length  
\* 29360 35962: contig of 6603 bp in length  
\* gap of unknown length  
\* 35963 43464: contig of 7502 bp in length  
\* gap of unknown length  
\* 43465 52200: contig of 8736 bp in length  
\* gap of unknown length  
\* 52201 74680: contig of 22480 bp in length  
\* gap of unknown length  
\* 74681 105370: contig of 30690 bp in length  
\* gap of unknown length  
\* 105371 144977: contig of 39607 bp in length  
\* gap of unknown length  
\* 144978 199437: contig of 54460 bp in length  
\* gap of unknown length  
\* 199438 252026: contig of 52589 bp in length.

FEATURES  
source Location/Qualifiers  
1. 252026  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="hRPK\_93\_A\_1"  
/clone\_id="RPC1-11 Human Male BAC"

BASE COUNT 72821 a 52000 c 52107 g 73751 t 1347 others  
ORIGIN

Query Match 7.7%: Score 18; DB 35; Length 252026;  
Best Local-Similarity 100.0%: Pred. No. 7.2;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 93 TGAATCAGTGAATGGAAA 110  
DB 84605 TGAATCAGTGAATGGAAA 84622

Search completed: August 6, 1999, 11:40:51  
Job time: 6537 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:14 : Search time 3075.15 Seconds

(without alignments)  
150.739 Million cell updates/sec

Title: US-09-049-696-12

Perfect score: 235  
Sequence: 1 GACACCGACCAATTCCCGAG.....CAGTGTAAAGTGGGGCTC 235

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST.\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*

54: em\_est23:\*  
55: em\_est24:\*  
56: em\_est25:\*  
57: em\_est26:\*  
58: em\_est27:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	176.4	75.1	479	35	AA541829	AA541829 vj01f01.r
2	176.4	75.1	450	37	AA734088	AA734088 vs19c08.r
3	113.2	48.2	470	43	AI197242	AI197242 u681f01.r
4	108.8	46.3	469	36	AA611564	AA611564 v085d05.r
5	65.8	28.0	420	49	AI641868	AI641868 vq39a03.y
6	59.6	25.4	524	39	AA871630	AA871630 vq39a03.r
7	52.6	22.4	449	38	AA759816	AA759816 vt81b01.r
8	47	20.0	494	33	AA430264	AA430264 zw65h07.r
9	47	20.0	502	33	AA446075	AA446075 zw66f06.r
10	34.2	14.6	422	47	AI504323	AI504323 v105h03.x
11	33.2	14.1	391	23	R86653	R86653 ym8b04.r1
12	31	13.2	467	24	H67652	H67652 yf76a01.s1
13	30.8	13.1	465	41	AI052319	AI052319 oy93g08.x
14	30.6	13.0	450	22	H05062	H05062 y169f02.s1
15	30	12.8	547	43	AI231949	AI231949 EST228637
16	30	12.8	473	45	AI388505	AI388505 q982f10.x
17	30	12.8	747	45	AI377123	AI377123 tc38c05.x
18	29.6	12.6	360	36	C61275	C61275 C61275 yu1j
19	29.6	12.6	367	42	AI151982	AI151982 udl1b02.r
20	29.6	12.6	585	43	AI166890	AI166890 xy1em.est
21	29.6	12.6	503	47	AI479306	AI479306 tm27a02.x
22	29.4	12.5	593	23	H18075	H18075 ym38f11.s1
23	29.4	12.5	605	40	AA939690	AA939690 v284b02.r
24	29.4	12.5	478	49	AU052130	AU052130 AU052130
25	29.2	12.4	389	21	R05588	R05588 ye92a02.r1
26	29.2	12.4	521	26	W30901	W30901 zc65c10.r1
27	29.2	12.4	462	26	W88914	W88914 zh69b04.r1
28	29.2	12.4	493	27	AA011701	AA011701 z103a07.r
29	29.2	12.4	280	32	AA370051	AA370051 EST81614
30	29.2	12.4	579	34	AA495497	AA495497 c105 Zhou
31	29.2	12.4	517	37	AA724783	AA724783 a101h01.s
32	29.2	12.4	243	38	AA750498	AA750498 ISGS0074
33	29.2	12.4	541	44	AI261363	AI261363 q232f12.x
34	29.2	12.4	518	46	AI418228	AI418228 tf75h03.x
35	29.2	12.4	546	46	AI438489	AI438489 486006G10
36	29.2	12.4	612	49	AI654023	AI654023 ly04g11.x
37	29	12.3	455	39	AA871197	AA871197 vq32c11.r
38	28.8	12.3	488	21	R18191	R18191 yf98f12.r1
39	28.8	12.3	528	22	R24983	R24983 yq41g08.r1
40	28.8	12.3	481	36	AA614014	AA614014 n082a07.s
41	28.8	12.3	506	44	AI266265	AI266265 qx89f11.x
42	28.8	12.3	448	46	AI415182	AI415182 mc06e09.x
43	28.8	12.3	438	48	AI585756	AI585756 vs01g02.x
44	28.6	12.2	359	23	D53511	D53511 HUM111A12B
45	28.6	12.2	450	25	N65515	N65515 20555 Lambd

#### ALIGNMENTS

RESULT 1  
LOCUS AA541829 479 bp mRNA  
DEFINITION vj01f01.r1 Barstead mouse pooled organs MFLNB4 Mus musculus cDNA  
clone IMAGE:920473 5' similar to TR:G1184066 G1184066  
CALCIDU-ACTIVATED CHLORIDE CHANNEL. ; mRNA sequence.  
ACCESSION AA541829  
NTD g2288263

VERSION	AA541829.1	GI:2288263
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 479)	
TITLE	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Schellenberg,K., Steptoe,M., Tan,F., Underwood,R., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.	
JOURNAL	The WashU-HMI Mouse EST Project	
COMMENT	Unpublished (1996) On Sep 12, 1996 this sequence version replaced gi:1406841.	
FEATURES	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNLN ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. <a href="mailto:MGI:532689">MGI:532689</a> Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 368. Location/Qualifiers 1..479 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:920473" /clone_lib="Barstead mouse pooled organs MPLRB4" /sex="mixed" /tissue_type="pooled organs" /dev_stage="7 day" /lab_host="DH10B" /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGATCGAGTCGAGCGGAGCCGCCGCCCTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTGGATTGCGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bob Barstead."	
BASE COUNT	139 a 115 c 120 g 105 t	
ORIGIN		
Query Match	75.1%; Score 176.4; DB 35; Length 479;	
Best Local Similarity	84.6%; Pred. No. 3.3e-46;	
Matches 198; Conservative	0; Mismatches 36; Indels 0; Gaps 0;	
OY	2 ACACAGCAAAATTCOCACGCCTCTGCTGTTAACAATAATTCGCCACAGAGCCTCCC	61
Db	152 ACACAGGGAATATCCCCACCCCTGTAACAGTATGACAGCATGCGCCAAGGAGCTGCG	211
OY	62 CAATTCTCAGGGCCAGTGTACAGCCCTGATTGAATCACTGATGAGNAACAGTACTCT	121
Db	212 CTATCTCAGGGCCAGCGTCACAGCCTTATTTGAATCTGTGAATGGAATAAACAGTAAACC	271
OY	122 TGGAACTACTGATATATGAGCAGAGGTGCTGATGCTACTACTAGAATGACGGTGTCTACTCAA	181
Db	272 TGGAAATCTCGATATACGAGAGCAGGTGCGGATGCCACCAAGATGATGCTCTACTCAA	331
OY	182 GGTATTTCACAACCTATGACAGAGATGGATATACAGTGTAAAGGCGGGCTC	235
Db	332 GGTTTTCACAGCTTTTCATGCAAAATGATGATACAGCCTTTAAAATATAGGGCTC	385

[illegible]

Db	224	TGGAATTCATGGATTAACGAGCAGCGAGTGC	CGGATGCACCAACGAATGATGTC	CTACTGAA	283
0y	182	GGTATTTCACAACTATATGACACGAA	TGGTAGTATACAGTAA	AGTGGCGCTC	235
Db	284	GGTTTTTACACGCTTTGTGATGC	CAATGTGATACAGCGTTTAA	TAATGCGGCTC	337
RESULT	3				
LOCUS	Al197242				
DEFINITION	Al197242.1	470 bp	mRNA	EST	14-OCT-1998
ACCESSION	U081610.1	Soares mouse mammary gland	NMLMG	Mus musculus	CDNA clone
VERSION	IMAGE:1477273.5	' similar to SW:EDIC	BOVIN P54281	EPITHELIAL	
KEYWORDS	CHLORIDE CHANNEL PROTEIN	;;	mRNA sequence.		
SOURCE	Al197242				
ORGANISM	EST.				
REFERENCE	1	(bases 1 to 470)			
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Gessel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Phaisang,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.				
TITLE	The Mashu-HHMI Mouse EST Project				
JOURNAL	Unpublished (1996)				
COMMENT	On Aug 21, 1998 this sequence version replaced.				
FEATURES					
Source	Contact: Marra M/Mouse EST Project Washu-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LINT ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:925629 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 456. Location/Qualifiers 1..470 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:1477273" /clone_lib="Soares mouse mammary gland NMLMG" /sex="female (lactating)" /tissue_type="mammary gland" /lab_host="DH10B" /note="Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRTT3 vector. Library is normalized. Library was constructed by Bento Soares and W. Fatima Bonaldo."				
BASE COUNT	148 a	103 c	116 g	103 t	
ORIGIN					
Query Match	48.2%	Score 113.2;	DB 43;	Length 470;	
Best Local Similarity	68.3%;	Pred. NO. 4.3e-26;			
Matches 157;	Conservative 0;	Mismatches 73;	Indels 0;	Gaps 0;	
0y	3	CACGACCAATTCCTCCACCCCTCTG	TGATTTATCCAAATATTCGCAAG	AGCCTCCC	62
Db	28	CACACCCAGTACCTACCGCGATAT	GTGTATGTACGGGTCACACCAAG	GCTTTTGC	87
0y	63	AATTCACAGGCGCAATGTCACAG	CCCTGATGTAATCAGTAAATGGA	AAAAACAGTTACCTT	122

[illegible]

FEATURES  
source

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image:1096588"
/clone_lib="Birstead bowel MPRB9"
/tissue_type="bowel"
/dev_stage="8 weeks"
/lai_host="DH10B"
/notes="Vector: p1773D-Pac (Pharmacia) with a modified
polylinker. Site.1: EcoRI; Site.2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACGACAGCTGAAGGAGGAGGCGGCCCTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATGCAACCTCTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified p1773 vector.
Source irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob
Birstead."

```

RESULT	6
LOCUS	AA871630
DEFINITION	AA871630 524 bp mRNA EST 16-MAR-1996
ACCESSION	U939803.r1 Barstad boxel MPRB9 Mus musculus cDNA clone
VERSION	IMAGE:1096888.5 similar to SW:EDLC_BOVIN P54281 EPITHELIAL
NID	CHLONIDE CHANNEL PROTEIN ; mRNA sequence.
KEYWORDS	AA871630
SOURCE	92967075
ORGANISM	AA871630.1 GI:2967075
	EST.
	house mouse.
	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
	1 (bases 1 to 524)
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubug
	Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
	Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
	Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
	Waterson,R.
TITLE	The WashU-HMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1397799.

FEATURES  
source

Contact: Maria M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:6028220  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 489.

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1096588"
/clone_lib="Bairstead bowel MFLRB9"
/tissue_type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: EcoRI. Site 2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTGACATCTCAAGTGGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[TAATGGCATCTCTT], digested with Not I and cloned into the

```







Query Match	14.68;	Score	34.2;	DB	47;	Length	422;
Best Local Similarity	58.38;	Pred. No.	0.59;				
Matches	60;	Conservative	0;	Mismatches	43;	Indels	0;
						Gaps	0;

QY 13 TTCCCCAGCCCTGTAGTTATGCAATATTGCCAAGAGCCCTCCCAATTCTAGG 72  
| | | | | | | | | | | | | | | | | | | | | |  
Db 103 TCCCGAGCCTGTGAAGTATTAAGCACCCTTATCCCCAGGAGCGGCCACTTACACAGT 162

**Qy** 73 GCCAGTGTCAACAGCCCTGATTCATAGTAGTGAAATGGAAAAACAG 115  
||| ||| ||| | ||| ||| |||  
**Db** 163 GCCCGTTCAAGGCCGAACACGAGCAATGCATGTACCAACAG 205

RESULT	11
R86653/c	
LOCUS	R86653 391 bp mRNA
DEFINITION	ym866504.r1 Soares adult brain N2b4H855Y Homo sapiens cDNA clone IMAGE:165775 5', mRNA sequence.

REFERENCE  
AUTHORS  
Dukerjyot; Metazoa; Chordata; Craniata; Vertebrata; Mammalia  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 391)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	On May 8, 1995 this sequence version replaced gi:800514.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewartson.wustl.edu  
Insert Size: 1493  
High quality sequence stops: 329 Source: IMAGE Consortium, LNLINK  
This clone is available royalty-free through LNLINK; contact the  
IMAGE Consortium ([info@image.lln.gov](mailto:info@image.lln.gov)) for further information.  
Insert Length: 1493 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 329.

## FEATURES

Source

/organism="Homo sapiens"  
 /db.xref="GDB:566453"  
 /db.xref="taxon:9606"  
 /clone\_image="IMAGE:165775"  
 /clone\_lib="Scotars adult brain N2b4HB55Y"  
 /sex="Male"  
 /dev\_stage="55-year old"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: brain; Vector: pTR73D (Pharmacia) with a  
 modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTCACATCTGAATGGGAGCGCGCCGCTTTTCTTTTCTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pTR73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 53. Library constructed by Bento  
 Soares and M. Fatima Bonaldo. The adult brain RNA was  
 provided by Dr. Donald B. Gilden. Tissue was acquired  
 17-18 hours after death which occurred in consequence of a  
 ruptured aortic aneurysm. RNA was prepared from a pool of

BASE COUNT	74 a	101 c	114 g	94 t	8 others
ORIGIN	tissues representing the following areas of the brain: frontal, parietal, temporal and occipital cortex from the left and right hemispheres, subcortical white matter, basal ganglia, thalamus, cerebellum, midbrain, pons and medulla.				

Query Match	14.1%	Score	33.2;	DB	23;	Length	391;
Best Local Similarity	51.1%;	Pred. No.	1.2;				
Matches	71;	Conservative	0;	Mismatches	68;	Indels	0;
				Gaps	0;		

QY	35	ATGCAAAATATTCGCCAAGGAGGACCCCTCCCAATTCACAGGGCCAGTGTACAGCCCTGATTTG	94
Db	323	AAGAAAATCTCAATTGTGGCCCTCACCCTCCCATGACACAGGAACTTTGTCAACGGCATTC	264
QY	95	AATCAGATGAATGAAAAACAGTTACCTTGGAACACTACTGATTAATGAGCAGGTCCTGATG	154
Db	263	GNCGTGTAATTCGMAACCAACCAACAGTCTCTGACGGGAAACCTCGAGCAGGACCAACAG	204
QY	155	CTACTAAGGATGACGGTGT	173
Db	203	CAACGAGGATGGGGGGGT	185

RESULT	12
H67652	
LOCUS	H67652 467 bp mRNA EST 18-OCT-1995
DEFINITION	y176601.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:	211176 3', mRNA sequence.

ORGANISM    Homo sapiens

REFERENCE  
AUTHORS

1 (bases 1 to 467)

Hillier, L., Lennon, G., Becker, M., Ronaldo, M. F., Chiappelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mards, E., Moore, B., Morris, M., Parsons, J., Prings, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J.,  
Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

TITLE	Generation and analysis of 280,000 human expressed sequence tags
JOURNAL	Genome Res. 6 (9), 807-828 (1996)
MEDLINE	97044478
COMMENT	On Nov 29, 1993 this sequence version replaced gi:501870.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 989  
High quality sequence stops: 289  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 989 Std Error: 0.00  
Seq Primer: Promega -21ml3  
High quality sequence stop: 289.

## FEATURES

### Source

```

/organism="Homo sapiens"
/db_xref="GDB:3784017"
/db_xref="taxon:9606"
/clone IMAGE:211176"
/clone_11b="Soares fetal liver spleen IMFLS

```

```

Robert Strausberg, Ph.D.
1) 496-1550
robert.strausberg@nih.gov
There is available royalty-free through LNL; contact the
consortium (info@medline.lnl.gov) for further information.
length: 1081 Std Error: 0.00
ers: 40m13 fwd. Et from Amersham
lity sequence stop: 372.
Location/Qualifiers
1..465
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="4p16.3"
/clone="IMAGE:1673438"
/clone_1lb="Soares_fetal_liver_spleen_INPIS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
on a modified polylinker Site_1: Pac I; Site_2: Eco RI"

```

RESULT	14
LOCUS	H05062/c
DEFINITION	H05062 450 bp mRNA EST 21-JUN-1995 Y169J02.s1 Soares Infant brain INIB Homo sapiens cDNA clone
ACCESSION	H05062 IMAGE:43294 3', mRNA sequence.
NID	9868614
VERSION	H05062.1 GI:868614
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1. (bases 1 to 450) Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapelli,B., Chisoe,S., Dietrich,N., Dubugue,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Marlis,E., Moore,B., Morris,M., Parsons,T., Prange,C., Rifkin,L., Rouffling,T., Scheilenderg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE	Generation and analysis of 280,000 human expressed sequence tags
JOURNAL	Genome Res. 6 (9), 807-828 (1996)
MEDLINE	97044478
COMMENT	

Wilson RK  
University School of Medicine  
West Park Parkway, Box 8501, St. Louis, MO 63108  
286 1800  
286 1810  
stewartson.wustl.edu  
size: 2167  
library sequence strops: 340  
IMAGE Consortium, LLNL  
name is available royalty-free through LLNL ; contact the  
consortium (info@image.llnl.gov) for further information.  
length: 2167 Std Error: 0.00  
er: Promega -21ml3  
library sequence strops: 355.  
location/Qualifiers  
1. 450  
/organism="Homo sapiens"  
/db\_xref="GDB:415835"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:43294"  
/clone\_lib="Scores Infant brain INTB"  
/sex="female"



---

*This Page Blank (uspto)*

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:40:47 ; Search time 650.92 Seconds  
(without alignments)  
90.326 Million cell updates/sec

Title: US-09-049-696-12

Perfect score: 235

Sequence: 1 GACACCCAGCAATTCCCGAG.....CAGGTAAAGTGGCGGCTC 235

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N.Geneseq\_36.\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	65	27.7	878	1	T45884	Human colon specif
2	65	27.7	878	1	V16672	Polynucleotide seq
3	17	7.2	2404	1	T33220	Oestrogen-regulate
4	17	7.2	2404	1	T99071	Partial sequence o
5	16	6.8	5091	1	O06164	The iga gene, enco
6	16	6.8	2515	1	O75304	Unique region of h
7	16	6.8	6828	1	O75302	Genomic RNA of hum
8	16	6.8	3134	1	T62135	Leishmania major M
9	16	6.8	5300	1	T90447	Thermotoga bacter
10	16	6.8	5102	1	V10261	Human COL4A6 gene.
11	16	6.8	3134	1	V47555	Leishmania antigen
12	16	6.8	7070	1	V58230	Omega-cyclohexane
13	16	6.8	5190	1	X13332	Enterococcus faeca
14	16	6.8	717	1	X20662	Polynucleotide seq
15	15	6.4	8298	1	N92777	CDNA sequence enco
16	15	6.4	8298	1	O50168	p190 protein gene.
17	15	6.4	3706	1	O54693	Another specific ge
18	15	6.4	110	1	O56834	Primate calicivirus
19	15	6.4	3706	1	O58340	Tobacco Ant32 geno
20	15	6.4	2747	1	O87915	Human histamine H1
21	15	6.4	483	1	T04236	CDNA up to the fou
22	15	6.4	1742	1	T04227	Human histamine H1
23	15	6.4	43	1	T04232	PNTV3606 amplifica
24	15	6.4	53577	1	T18551	Human polycystic k
25	15	6.4	38	1	V16077	PCR primer used to
26	15	6.4	53577	1	T94108	Human PRD1 locus b
27	15	6.4	53526	1	T94101	Human PRD1 gene. H
28	15	6.4	9472	1	V25601	Feline immunodefic
29	15	6.4	3879	1	V21060	CDNA encoding the
30	15	6.4	890	1	T84051	DNA encoding S. au
31	15	6.4	3879	1	V29009	Human polynucleot
32	15	6.4	4727	1	V26975	Consensus VCP255 H
33	15	6.4	4616	1	V26971	Feline immunodefic
34	15	6.4	2032	1	V26974	Consensus VCP255 H
35	15	6.4	4150	1	V26976	Consensus VCP255 H
36	15	6.4	5319	1	V35235	R. prowazekii S-1a
37	15	6.4	17710	1	V31256	E. coli J96 pathog
38	15	6.4	735	1	V52499	Streptococcus pneu
39	15	6.4	3763	1	V52319	Streptococcus pneu
40	15	6.4	10357	1	V52324	Streptococcus pneu
41	15	6.4	2393	1	V52223	Streptococcus pneu
42	15	6.4	11964	1	V52194	Streptococcus pneu
43	15	6.4	2674	1	V42984	Streptococcus pneu

## ALIGNMENTS

c	44	15	6.4	110000	1	V21209_11	Continuation (12 o
	45	15	6.4	1299	1	X30776	Streptococcus pneu
ALIGNMENTS							
RESULT	1						
ID	T45884	T45884 standard; CDNA: 878 BP.					
AC	T45884;						
DT	13-MAR-1997	(first entry)					
DE	Human colon specific gene CSG5 CDNA partial clone.						
KW	Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;						
OS	Human sapiens.						
FH	Key	Location/Qualifiers					
FT	cds	2..691					
FT	cds	/*tag= a					
PN	W09639419-A1.						
PD	12-DEC-1996.						
PF	06-JUN-1995; U07289.						
PR	06-JUN-1995; W0-U07289.						
PA	(HUMA-) HUMAN GENOME SCI INC.						
PI	Rosen CA, Xu G;						
PI	WPI: 97-043054/04.						
DR	P-PSDB: W06548.						
PT	Human colon specific genes and their expression products - detection						
PT	of which, in non-colon tissue samples, can be used as indication of						
PT	colon cancer metastasis						
PS	Claim 1; Fig 5; 60pp; English.						
CC	13 CDNA clones (T45880-92), most of them partial clones, correspond						
CC	to human colon specific genes, designated CSG1, CSG2, etc., that						
CC	are primarily expressed in tissues derived from the colon. CSG7						
CC	and CSG10 show reduced expression in colon cancer cells as compared						
CC	to that in normal cells; the remaining genes are overexpressed in						
CC	colon cancer. The partial cDNA sequences can be used to isolate						
CC	full-length clones and genomic clones including the complete gene.						
CC	CSG nucleic acids can be used to produce CSG polypeptides (see also						
CC	W06545-53) in transformed host cells, as probes to detect disorders						
CC	of the colon, partic. colon cancer and colon cancer metastasis, and						
CC	in gene therapy.						
SO	Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;						
Query Match							
Best Local Similarity 100.0%; Score 65; DB 1; Length 878;							
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
QY	171	TGCTACTCAAGGATTTTCACACTTATGACAGGATGTAGTACGTTAAAGTCG	230				
DB	1	TGCTACTCAAGGATTTTCACACTTATGACAGGATGTAGTACGTTAAAGTCG	60				
QY	231	GGCTC 235					
DB	61	GGCTC 65					
RESULT 2							
ID	V16672	V16672 standard; CDNA: 878 BP.					
AC	V16672;						
DT	22-JUN-1998	(first entry)					
DE	Polynucleotide sequence of a colon-specific gene.						
KW	Colon-specific gene; probe; detection; expression; human;						
OS	Human sapiens.						
FH	Key	Location/Qualifiers					
FT	CDS	2..685					
FT	CDS	/*tag= a					
PN	US5733748-A.						
PD	31-MAR-1998.						

PF 06-JUN-1995; 469667.  
 PR 06-JUN-1995; US-469667.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen C, Yu G;  
 DR WPI: 98-229823/20.  
 DR P-PSDB: W46879.  
 PT Colon-specific nucleic acids - useful as probes for detecting colon  
 PT cancer micrometastases  
 PS Claim 15; Fig 5A-B; 51pp; English.  
 CC V16668-81 represent polynucleotide sequences of partial or full-length  
 CC cDNA clones of colon-specific genes. The polynucleotides can be used  
 CC as probes to detect expression of the corresponding human genes, e.g. in  
 CC diagnostic assays for detecting micrometastases of colon cancer.  
 CC Recombinant cells containing the polynucleotides can be used to  
 CC produce the polypeptides, in order that antibodies can be raised and  
 CC used in further screening or diagnostics.  
 SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 27.7%; Score 65; DB 1; Length 878;  
 Best Local Similarity 100.0%; Pred. No. 2e-25;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 171 TGTCTACTGAGTATTTCACACTATGACAGCATGATGATGACGTAAGTGGC 230  
 DB 1 TGTCTACTGAGTATTTCACACTATGACAGCATGATGATGACGTAAGTGGC 60  
 OY 231 GGCTC 235  
 DB 61 GGCTC 65

RESULT 3  
 T33220/c  
 ID T33220 standard; cDNA; 2404 BP.  
 AC T33220;  
 DT 16-OCT-1996 (first entry)  
 DE Oestrogen-regulated p171 gene partial 3' extended cDNA clone.  
 KW Oestrogen-regulated gene; p171; breast cancer; tumour; diagnosis;  
 KW marker; metastasis; ss.  
 OS Homo sapiens.  
 FH Key  
 FT cds  
 FT 1..1296  
 FT /\*tag- a  
 FT 1297..2404  
 FT /\*tag- b  
 FT misc\_difference 2201..2260  
 FT /\*tag- c  
 FT /note- "bases 2201-2260 are identical to bases  
 FT 2141-2200"  
 FT 2141-2200"  
 PN CA2132500-A.  
 PD 21-MAR-1996.  
 PR 20-SEP-1994; 132500.  
 PR 20-SEP-1994; CA-132500.  
 PA (UYMA-) UNIV WALES COLLEGE OF MEDICINE.  
 PI Gee JM, Green CD, Manning DL, Nicholson RI;  
 DR WPI: 96-268987/28.  
 DR P-PSDB: R88004.  
 PT Oestrogen-regulated p171 gene and deduced polypeptide - useful for  
 PT predicting the propensity for metastatic spread or the  
 PT responsiveness to endocrine treatment of breast tumour  
 PS Disclosure: Page 16-17; 27pp; English.  
 CC A cDNA clone (T33219) corresponding to a portion of the oestrogen-  
 CC regulated human p171 gene was extended (T33220) using a primer  
 CC directed cloning strategy. p171 is a candidate for the hormonal  
 CC regulation of tumour invasion of breast cancer, a prognostic marker  
 CC of metastatic spread, and a suitable target for therapeutic  
 CC intervention by antihormones and antibody-directed methods. It is  
 CC significantly associated with oestrogen-dependent (ER+) disease.  
 CC The polypeptide (R88004) encoded by the cDNA contains 3 immunogenic  
 CC regions.  
 SQ Sequence 2404 BP; 726 A; 413 C; 490 G; 775 T;

Query Match 7.2%; Score 17; DB 1; Length 2404;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 GCCCTCCCAATTCACG 71  
 DB 1748 GCCCTCCCAATTCACG 1732

RESULT 4  
 T39071/c  
 ID T39071 standard; cDNA; 2404 BP.  
 AC T39071;  
 DT 24-MAR-1998 (first entry)  
 DE Partial sequence of the p171 gene.  
 KW Oestrogen regulated gene; p171 gene partial sequence; breast cancer;  
 KW metastatic risk assessment; hormone therapy responsiveness;  
 KW tumour oestrogen receptor expression; lymph node involvement; ds.  
 OS Homo sapiens.  
 FH Key  
 FT cds  
 FT 1..1296  
 FT /\*tag- a

PN US5693465-A.  
 PD 02-DEC-1997.  
 PR 22-SEP-1994; 311023.  
 PR 22-SEP-1994; US-311023.  
 PA (UYMA-) UNIV WALES COLLEGE OF MEDICINE.  
 PI Gee JM, Green CD, Manning DL, Nicholson RI;  
 DR WPI: 98-031740/03.  
 DR P-PSDB: W34528.  
 PT Assessment of metastatic risk or oestrogen responsiveness in breast  
 PT cancer - by detecting expression of p171 protein  
 PS Disclosure: Column 11-14; 11pp; English.  
 CC This sequence represents a partial sequence of the p171 gene. This  
 CC sequence extends further in the 3' direction than the p171 partial  
 CC sequence shown in T39070. The protein encoded by this sequence is an  
 CC oestrogen regulated gene associated with breast cancer. The protein  
 CC encoded by this sequence can be used in the method of the invention. The  
 CC method is for determining the risk of metastasis of a female breast  
 CC tumour and predicting the responsiveness of a female breast  
 CC tumour to hormone treatment comprising determining if a tissue sample from the tumour  
 CC expresses a polypeptide comprising at least 14 continuous amino acids of  
 CC the protein encoded by this sequence. The method is useful for  
 CC determining the risk of metastasis of a female breast tumour and to  
 CC predict the response of the female breast tumour to hormone treatment.  
 CC p171 expression is highly predictive of tumour oestrogen receptor  
 CC expression and lymph node involvement.  
 SQ Sequence 2404 BP; 726 A; 411 C; 492 G; 775 T;

Query Match 7.2%; Score 17; DB 1; Length 2404;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 GCCCTCCCAATTCACG 71  
 DB 1748 GCCCTCCCAATTCACG 1732

RESULT 5  
 006164  
 ID 006164 standard; DNA; 5091 BP.  
 AC 006164;  
 DT 31-JAN-1991 (first entry)  
 DE The iga gene, encoding IgA1 protease.  
 KW IgA1; vaccine; meningitis; gonorrhoea; allergies; ss.  
 OS Hemophilus influenzae.  
 FH Key  
 FT cds  
 FT 262..4887  
 FT /\*tag- a  
 FT /product-IgA1 protease  
 FT -10\_signal  
 FT 235..240

```

FT      -35_signal      /*tag= b
FT      211..215      /*tag= c
FT      rbs      248..253
FT      /*tag= d
PN      MO9011367-A..
PD      04-OCT-1990.
PF      16-MAR-1990; DK0073.
PR      17-MAR-1989; DK-001308.
PA      (KILI/) KILIAN M.
PI      Kilian M, Poulsen K;
DR      WPI: 90-320267/42.
DR      P-PSDB: R07304.
PT      Immunoglobulin A1 protease prodn. - by cloning from
PT      microorganisms for immunisation against immunoglobulin A1
PT      protease producing bacteria
PS      Disclosure: fig 3: 44pp; English.
CC      This Iga gene is from H. influenzae serotype b strain HK368. On
CC      transformation of E. coli cells with a vector contg. this gene,
CC      immunoglobulin (19)A1 protease is produced which is useful in a
CC      vaccine for e.g. meningococcal meningitis, gonorrhoea or allergic
CC      diseases.
SQ      Sequence 5091 BP; 1862 A; 873 C; 979 G; 1377 T;

```

```

Query Match      6.8%; Score 16; DB 1; Length 5091;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      101 TGAATGAAAAACAGT 116
DB      701 TGAATGAAAAACAGT 716

```

```

RESULT 6
ID      075304/c
ID      075304 standard; Dna: 2515 BP.
AC      075304;
DR      06-AUG-1995 (first entry)
DE      Unique region of human astrovirus serotype 2 subgenomic RNA.
KW      Astrovirus; serotype 2; H-Ast 2; gastroenteritis; antigen; ss.
OS      Human astrovirus serotype 2.
PN      MO9426902-A.
PD      24-NOV-1994.
PF      12-MAY-1994; U05287.
PR      12-MAY-1993; US-061465.
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI      Glass RI, Jlang B, Koopmans M, Monroe SS;
DR      WPI: 95-006798/01.
PT      New nucleic acid from human astrovirus serotype 2 - and related
PT      antigenic polypeptide(s) and antibodies, useful in diagnosing
PT      infections and prodn. of vaccines
PS      Claim 17: Page 42-43; 59pp; English.
CC      The consensus sequence for the unique region of the subgenomic RNA
CC      of human astrovirus serotype 2 is given in 075304. It comprises an
CC      1-nt 5' UTR; a 2388-nt ORF; and an 85 3' UTR (GenBank Data Library
CC      L06802). The single ORF encodes a 796 AA polypeptide with a
CC      predicted Mol. Mass of 88 kDa. At the amino terminus, the predicted
CC      polypeptide has a region of basic AAs that may play a role as a
CC      nucleic acid binding motif. At the carboxy terminus is a region of
CC      acidic AAs. See also 075302 CC on ORF 2.
SQ      Sequence 2515 BP; 787 A; 558 C; 567 G; 603 U;

```

```

Query Match      6.8%; Score 16; DB 1; Length 2515;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      16 CCCAGCCCTCTGTAG 31
DB      1830 CCCAGCCCTCTGTAG 1815

```

```

RESULT 7
ID      075302/c
ID      075302 standard; RNA: 6828 BP.
AC      075302;
DR      06-AUG-1995 (first entry)
DE      Genomic RNA of human astrovirus serotype 2.
KW      Astrovirus; serotype 2; H-Ast 2; gastroenteritis; antigen; ss.
OS      Human astrovirus serotype 2.
FH      Key      Location/Qualifiers
FT      cds      83..2842
FT      /*tag= a
FT      1a      /label= ORF 1a
FT      2773..4329
FT      /*tag= b
FT      1b      /label= ORF 1b
FT      4325..6712
FT      /*tag= c
FT      2      /label= ORF 2

```

```

PN      MO9426902-A.
PD      24-NOV-1994.
PF      12-MAY-1994; U05287.
PR      12-MAY-1993; US-061465.
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI      Glass RI, Jlang B, Koopmans M, Monroe SS;
DR      WPI: 95-006798/01.
DR      P-PSDB: R67358.
PT      New nucleic acid from human astrovirus serotype 2 - and related
PT      antigenic polypeptide(s) and antibodies, useful in diagnosing
PT      infections and prodn. of vaccines
PS      Claim 1; Page 34-38; 59pp; English.
CC      Genomic RNA of H-Ast 2 is 6797 nts in length excluding the polyA
CC      tail (31 As). It is deposited (ATCC L13745). It has three
CC      overlapping reading frames (ORFs) (see FT). The 5' UTR which
CC      precedes ORF 1a is predicted to contain extensive secondary
CC      structure, as demonstrated by the characteristic stem-loop
CC      structures preceding the initiation AUG codon. ORF 1b overlaps ORF
CC      1a by 70 nts. It is in reading frame +1. Its first AUG codon is
CC      predicted to be weak. ORF 2, present also in the subgenomic RNA,
CC      overlaps ORF 1b by 5 nts. It encodes a capsid protein precursor of
CC      796 AAs with a predicted MW of 88 kDa. H-Ast 2 genomic RNA is claimed
CC      and so are the nucleic acid sequences and antigens of ORFs 1a, 1b
CC      and 2. When the overlap region of ORF 1a and 1b was examined, a
CC      ribosomal frameshift signal was identified, consisting of the
CC      'shift' heptanucleotide (AAAAAC) from posn. 2791-2797, followed
CC      by the stem-loop structure that may form a pseudoknot with a
CC      downstream sequence. The ribosomal frameshifting during
CC      translation of astrovirus RNA directs the synthesis of an ORF
CC      1a/1b fusion nonstructural polyprotein of 1416 AAs with a predicted
CC      Mol. mass of 161 kDa (see R67358). The predicted transmembrane
CC      alpha-helices occur at residues 156-172, 308-333, 343-362, and
CC      369-387, the predicted cleavage site at the N-terminus of the
CC      putative VPg-protease occurs at residues 419-420, the putative
CC      nuclear localisation signal occurs at residues 666-682, and fusion
CC      dipeptide (KK) occurs at residues 904-905. The ORF 1a/1b can be
CC      identified at nts 83-4329 of 075302. The minus 1 frameshift occurs
CC      at posn. 2794 of the sequence. An isolated nucleic acid encoding
CC      ORF 1a/1b is given in 075303 where the minus 1 frame shift occurs
CC      at posn. 2712. Antigen 1a/1b is given in R67358.
SQ      Sequence 6828 BP; 2106 A; 1448 C; 1518 G; 1756 U;

```

```

Query Match      6.8%; Score 16; DB 1; Length 6828;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      16 CCCAGCCCTCTGTAG 31
DB      6143 CCCAGCCCTCTGTAG 6128

```

```

RESULT 8
ID      T62135/c
ID      T62135 standard; cDNA: 3134 BP.

```

AC	T62135;
AT	13-JUL-1997 (first entry)
DE	Leishmania major M15 antigen cDNA.
KW	Leishmaniasis; vaccine; diagnosis; M15; antigen; immunogen; ss.
OS	Leishmania major strain Friedlan.
FH	Key Location/Qualifiers
FT	cds
FT	421..2061
FT	/tag= a
PN	W09711180-A1.
PD	27-MAR-1997.
PF	20-SEP-1996; U15185.
PR	22-SEP-1995; US-533669.
PA	(CORI-) CORIXA CORP.
PI	Campes-Neto A. Dillon DC, Reed SG, Skeiky YA, Webb JR:
DR	WPI: 97-202888/18.
P-PSDB:	W11214.
PT	Polypeptide(s) comprising immunogenic part of Leishmania antigen -
PT	useful for leishmaniasis vaccination and diagnosis
PS	Claim 4: Page 42-47: 106pp: English.
CC	A CDNA clone (T62135), designated pfl-1, codes for the M15 antigen
CC	(W11214) of Leishmania major Friedlan strain. An L. major
CC	amastigote cDNA expression library prepd. In lambda ZAPII was
CC	screened using sera obtd. from infected BALB/c mice. A 5' fragment
CC	of an isolated partial clone was used to rescreen the expression
CC	library, and the clone with the largest insert (pfl-1) was
CC	selected. DNA molecules (see also T62136-39) encoding Leishmania
CC	antigens can be incorporated into vectors and used to produce
CC	recombinant antigens, or immunogenic portions of them, in host
CC	cells for use in vaccines against Leishmaniasis, or may themselves
CC	be used in vaccines for in vivo immunogen generation.
SQ	Sequence 3134 BP; 758 A; 843 C; 866 G; 663 T;

  

Query Match	6.8%; Score 16; DB 1; Length 3134;
Best Local Similarity	100.0%; Pred. No. 21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

  

QY	162 GGATGACGGTGCTCAG 177
Db	904 GGATGACGGTGCTCAG 889

  

```

RESULT      9
T90447      T90447 standard; DNA: 5300 BP.
ID          AC T90447;
DE          27-JAN-1998 (first entry)
DT          Thermoaerobacter thermohydrosulfuricus DNA polymerase gene.
KW          DNA polymerase; exonuclease; strand displacement amplification;
            KW Thermoaerobacter thermohydrosulfuricus; reverse transcription;
            KM cDNA; ss.
OS          Thermoaerobacter thermohydrosulfuricus.
FH          Key Location/Qualifiers
FT          CDS 1056..3674
FT          /tag= a
PN          W09721821-A1.
PD          19-JUN-1997.
PE          13-DEC-1996; U20225.
PR          15-DEC-1995; US-008688.
PA          (AMSH ) AMERSHAM LIFE SCI INC.
PI          Davis M, Mamone JA, Sha D;
DR          WPI: 97-332792/30.
DR          P-PSDB: W26605.
PT          New DNA polymerase from Thermoaerobacter thermohydrosulfuricus -
PT          useful for strand displacement amplification, reverse transcription
PT          and sequencing
PS          Example 1: Page 48-49; 92pp; English.
CC          This DNA sequence encodes a novel thermostable DNA polymerase
CC          (W26605) of the thermophilic anaerobe Thermoaerobacter
CC          thermohydrosulfuricus. It was isolated from a subgenomic library
CC          using a DNA polymerase I-specific probe that had itself been
CC          generated using primers (see T90452-53) based on conserved regions
CC          of DNA polymerase I family members. A host cell comprising a
    
```

```
CC vector containing an isolated nucleic acid encoding DNA polymerase  
CC is claimed. The DNA polymerase can be used in claimed methods: (a)  
CC for strand displacement amplification (SDA) when used with a  
CC thermostable restriction enzyme; (b) for production of cDNA when  
CC used in conjunction with polymerase and deoxyribonucleotides; and  
CC (c) to provide reverse transcription activity in RT-PCR. Mutagenic  
CC methods have been used to generate exonuclease-deficient DNA  
CC polymerases (see W2606-07) suited for sequencing, SDA and CDNA  
CC preparation.  
SQ Sequence 5300 BP; 2136 A; 587 C; 1036 G; 1541 T;  
  
Query Match 6.8%; Score 16; DB 1; Length 5300;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0  
  
QY 104 ATGGAAAAACACTTAC 119  
|||||  
DB 4232 ATGGAAAAACAGTTAC 4247  
  
RESULT 10  
V10261  
ID V10261 standard; DNA; 5102 BP.  
AC V10261.  
DT 03-JUN-1998 (first entry)  
DE Human COL4A6 gene.  
KW COL4A6 gene; type IV collagen protein; alpha-6(IV); Alports syndrome;  
KM leiomyomatosis; diagnosis; gene therapy; antibody; ds.  
KN Homo sapiens.  
FH Key  
FD Location/Qualifiers  
CDS  
2..5102  
/*tag_a  
/product= COL4A6 gene  
/transl_except= (pos: 83..85, aa: Xaa)  
/transl_except= (pos: 98..100, aa: Xaa)  
/transl_except= (pos: 4400..4402, aa: Xaa)  
/transl_except= (pos: 4421..4423, aa: Xaa)  
/transl_except= (pos: 4466..4468, aa: Xaa)  
/transl_except= (pos: 4877..4879, aa: Xaa)  
/note= "partial sequence of COL4A6 gene. This sequence  
contains various stop codons which are not  
represented in the translated protein sequence  
of W40109"  
  
US5731192-A.  
PN 24-MAR-1998.  
PD 23-JUN-1995; 494168.  
PR 27-AUG-1993; US-112465.  
PA (UYUA ) UNIV YALE.  
PI Readers SM, Zhou J;  
DR WPT: 98-216485/19.  
P-PSDB: W40109.  
FT Nucleic acid encoding human alpha-6(IV) collagen - useful for, e.g.  
FT diagnosis or gene therapy of Alport's disease  
FS Claim 2; Fig 1; 43pp; English.  
CC This sequence is a novel human COL4A6 gene which encodes an alpha-6(IV)  
CC collagen protein. This protein can be used for diagnosis or gene therapy  
CC of diseases associated with collagen type IV pathology, especially  
CC Alport's syndrome and associated diffuse leiomyomatosis. The polypeptide  
CC may also be used for generating monoclonal or polyclonal antibodies  
CC having specificity for the alpha-6(IV) polypeptide especially an antibody  
CC that is not crossreactive with other collagen proteins including  
CC alpha-1(IV), alpha-2(IV) and alpha-5(IV) collagens. This nucleotide  
CC sequence contains stop codons which do not appear in the encoded protein  
CC represented in W40109.  
SQ Sequence 5102 BP; 1188 A; 1356 C; 1540 G; 1018 T;  
  
Query Match 6.8%; Score 16; DB 1; Length 5102;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



Db 949 CCTTGACTACTGGA 964

## RESULT 11

VA7555/C  
VA7555 standard; DNA: 3134 BP.

AC VA7555:

DE 13-NOV-1998 (first entry)

Leishmania antigen M15 coding sequence.

Leishmania antigen; immune response; infection detection; therapy;

humoral response induction; cellular response induction; cancer;

interleukin-12 production; ss.

OS Leishmania sp.

FT Key Location/Qualifiers

FT CDS 421..2061

FT /tag- a

FT /product- M15

PD WO9835045-A2.

PD 13-AUG-1998.

PF 12-FEB-1998; U03002.

PR 27-AUG-1997; US-920609.

PR 12-FEB-1997; US-798841.

PA (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Reed SG, Skeiky YA, Webb JR;

DR P-PSDB: W70209.

PT New immunogenic fragments of Leishmania antigens and related nucleic acid, vectors and host cells - are useful for diagnosis, prevention

PT and treatment of leishmaniasis, also to induce production of

PT interleukin-12 generally

PS Claim 4; Page 66-69; 134pp; English.

CC This sequence encodes a Leishmania antigen (LAg) of the invention,

CC designated M15. Compositions and vaccines containing the protein are

CC used to generate a protective or therapeutic immune response against the

CC Leishmania species donavani, chagasi, infantum, major, amazonensis,

CC braziliensis, panamensis, tropica or guyanensis. They can also be used

CC to detect infection (in a skin test). The compositions induce a humoral

CC and/or cellular response, specifically of Th1 type, particularly

CC including induction of interleukin-12 (IL-12) production. They may thus

CC be used more generally to treat any condition (e.g. bacterial, viral or

CC protozoal infection, or cancer) which responds to IL-12.

CC Sequence 3134 BP; 758 A; 843 C; 866 G; 663 T;

QY 162 GGATGACGGTCTTAC 177

DB 904 GGATGACGGTCTTAC 889

## RESULT 12

V58230 standard; DNA: 7070 BP.

AC V58230:

DE 26-NOV-1998 (first entry)

Omega-cyclohexane fatty acid biosynthesis enzymes #2 ORF1-3 encoding DNA.

Omega-cyclohexane fatty acid; biosynthesis; enzyme; detection;

Allycyclobacillus acidocaldarius ATCC 27009; identification; microbe; ds.

OS Allycyclobacillus acidocaldarius.

FT Key Location/Qualifiers

FT CDS 1..1269

FT /tag- a

FT /label- ORF1

FT 1422..2270

FT /tag- b

FT /label- ORF2

FT /transl\_except- (pos:1422..1424,aa:Met)

FT 3387..4349

FT /tag- c

FT /label- ORF3

FT /function- "Beta-ketoacyl (acyl carrier protein)

FT synthase"

PD J10234376-A.

PD 08-SEP-1998.

PF 28-FEB-1997; 046570.

PR 28-FEB-1997; JP-046570.

PA (KIRI) KIRIN BEVERAGE KK.

DR WPI: 98-535030/46.

P-PSDB: W71639, W71640, W71641.

PT New nucleic acid - useful for detection and identification of genus

PT Allycyclobacillus microorganism(s)

PS Claim 8; Page 18-23; 37pp; Japanese.

CC The present sequence represents a new nucleic acid which encodes enzymes

CC which participate in the biosynthesis of Omega-cyclohexane fatty acid.

CC The nucleic acid is isolated from Allycyclobacillus acidocaldarius. The

CC present invention also describes primers and probes containing all or

CC part of the nucleic acid from Allycyclobacillus acidocaldarius. The

CC primers and probes may be used for detection and/or identification of a

CC microorganism of genus Allycyclobacillus. The method can detect and

CC identify Allycyclobacillus genus rapidly and easily.

CC Sequence 7070 BP; 1397 A; 2322 C; 1999 G; 1352 T;

QY 145 GGATGACGGTCTTAC 160

DB 6777 GGATGACGGTCTTAC 6792

## RESULT 13

X13332 standard; DNA: 5190 BP.

ID X13332

AC X13332:

DE 19-MAR-1999 (first entry)

Enterococcus faecalis genome contig SEQ ID NO:395.

KW Enterococcus faecalis; contig; detection; Enterococcal infection;

OS Enterococcus faecalis.

PD WO9850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; U08985.

PR 14-NOV-1997; US-066009.

PR 06-MAY-1997; US-044031.

PR 16-MAY-1997; US-046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI: 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.

PS Claim 1; Page 1586-1588; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC X12938 to X13919 represent these nucleotide sequences which are primary

CC nucleotide sequences, also known as contigs. The computer-based system

CC can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosis of Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

CC Sequence 5190 BP; 1785 A; 588 C; 1020 G; 1719 T;

Query Match 6.8%; Score 16; DB 1; Length 5190;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 26 TCGTAGTTATGCAAA 41  
 DB 2596 TCGTAGTTATGCAAA 2611

RESULT 14  
 X20662  
 ID X20662 standard; DNA: 717 BP.  
 AC X20662;  
 DT 05-MAY-1999 (first entry)  
 DE Polynucleotide sequence from the genome of Treponema pallidum.  
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;  
 KM enzyme production; ds.  
 OS Treponema pallidum.  
 PN M09859034-A2.  
 PD 30-DEC-1998.  
 PE 23-JUN-1998; U13041.  
 PR 24-JUN-1997; US-050667.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Fraser CM;  
 DR WPI: 99-081273/07.  
 PT New isolated Treponema pallidum nucleic acids - used to develop  
 products for the detection, diagnosis, characterisation, prevention  
 and therapy of T. pallidum infections, particularly syphilis  
 PS Claim 1; Page 823-824; 1150pp; English.  
 CC X20500-21243 represent polynucleotide sequences from the genome of  
 Treponema pallidum. The sequences can be used for detection,  
 diagnosis, characterisation, prevention and therapy for T. pallidum  
 infections, particularly syphilis. They can also be used for detecting  
 diseases related to Borrelia infections in animals, and for the  
 production of biosynthetic products such as enzymes.  
 CC Sequence 717 BP; 202 A; 189 C; 196 G; 129 T;  
 SQ

Query Match 6.8%; Score 16; DB 1; Length 717;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 TCTCAGGCGCAGTCTC 81  
 DB 542 TCTCAGGCGCAGTCTC 557

RESULT 15  
 N92777  
 ID N92777 standard; DNA: 2009 BP.  
 AC N92777;  
 DT 23-MAY-1990 (first entry)  
 DE cDNA sequence encoding KCCEAI antigen.  
 KW Carcinoembryonic antigens (CEA's); antibody (anti-CEA) production;  
 KM cancer; ss.  
 FH Key Location/Qualifiers  
 FT polyA\_site 2005..2010  
 FT CDS /\*\*tag= a 85..1341  
 FT /\*\*tag= b  
 PN EP-346710-A.  
 PD 20-DEC-1989.  
 PE 03-JUN-1989; 110096.  
 PR 21-NOV-1988; US-274107,US-207678.  
 PA (MOLE-) Molecular Diagnosti.  
 PI Barnett TR, Eling JJ, Kamarek ME, Kretschnmer A;  
 DR WPI: 89-372000/51.  
 DR P-PSDB: P93397.  
 PT DNA's coding for members of carcinoembryonic antigen family - used to  
 prepare probes to detect antigen or in prodn. of polypeptide(s) and  
 PT antibodies.  
 PS Claim 1; Pages 61-62; 65pp; English.  
 CC This sequence encodes the KCCEAI peptide sequence. When the DNA

CC is inserted into an expression vector and a host cell transformed with  
 CC this the resultant KCCEAI antigen product is a useful tool in cancer  
 CC diagnosis. Antibodies against this antigen can be used to detect its  
 CC presence in animal and human patients and in conjunction with eg a  
 CC toxin these can destroy a KCCEAI - expressing cell.  
 SQ Sequence 2009 BP; 597 A; 531 C; 398 G; 482 T; 1 Others;

Query Match 6.4%; Score 15; DB 1; Length 2009;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 211 AGATACAGTGTAAAA 225  
 DB 1940 AGATACAGTGTAAAA 1954

Search completed: August 5, 1999, 17:40:51  
 Job time: 6286 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:42:36 ; Search time 3268.17 Seconds

(without alignments)  
235.494 Million cell updates/sec

Title: US-09-049-696-16

Perfect score: 242

Sequence: 1 GTTATTCCTCCACAGACTC.....ATGTAATTTAGACTTCCTGT 242

Scoring table: OLIGO\_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl:.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pl3.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_st.\*  
14: gb\_sts.\*  
15: gb\_sy.\*  
16: gb\_un.\*  
17: gb\_vl.\*  
18: em\_fun.\*  
19: em\_hlg.\*  
20: em\_hum1.\*  
21: em\_hum2.\*  
22: em\_in.\*  
23: em\_om.\*  
24: em\_or.\*  
25: em\_ov.\*  
26: em\_pat.\*  
27: em\_ph.\*  
28: em\_pl.\*  
29: em\_ro.\*  
30: em\_sts.\*  
31: em\_sy.\*  
32: em\_un.\*  
33: em\_vl.\*  
34: gb\_hlg1.\*  
35: gb\_hlg2.\*  
36: gb\_in1.\*  
37: gb\_in2.\*  
38: em\_ba1.\*  
39: em\_ba2.\*  
40: em\_hum3.\*  
41: em\_hum4.\*  
42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	213	88.0	2826	42	AF127036	Homo sapi

2	202	83.5	3311	11	AF039400	AF039400 Homo sapi
3	162	66.9	35278	11	AF039401	AF039401 Homo sapi
4	63	26.0	878	5	195746	195746 Sequence 8
5	23	9.5	2937	12	AB017156	AB017156 Mus muscu
6	21	8.7	94038	8	ATAC004401	AC004401 Arabidops
7	21	8.7	86686	8	ATAC004786	AC004786 Arabidops
8	21	8.7	179854	34	CEY19D2	AL021447 Caenorhabd
9	21	8.7	39973	36	CEM106	AB025615 Arabidops
10	20	8.3	83544	7	AB025615	AC002994 Homo sapi
11	20	8.3	143819	11	AC002994	AE001386 Plasmodiu
12	12	8.3	12940	37	AE001386	U63091 Oryctolagus
13	19	7.9	15609	3	OCU63091	Z83827 Human DNA s
14	19	7.9	135686	9	HS47306	AL008967 Human DNA
15	19	7.9	128169	9	HS47306	Z72519 Human DNA s
16	19	7.9	125787	10	HSJ1817A	AC006288 Homo sapi
17	19	7.9	220218	11	AC006288	AF001905 Homo sapi
18	19	7.9	197620	11	AF001905	AF068625 Mus muscu
19	19	7.9	4192	12	AF068625	G48586 SHGC-82343
20	19	7.9	701	14	G48586	AL031767 Homo sapi
21	19	7.9	176564	34	HS135L22	Z98051 Homo sapien
22	19	7.9	79760	34	HS501A4	U88180 Caenorhabdi
23	19	7.9	34995	36	CELT27A3	Z73911 Caenorhabdi
24	19	7.9	39220	36	CET12A7	U01006 Herpetomona
25	19	7.9	547	36	U01006	U01007 Herpetomona
26	19	7.9	547	36	U01007	AC004722 Drosophi1
27	19	7.9	76619	37	AC004722	AF011388 Brevibacti
28	19	7.9	47202	37	AC004724	AB014462 Xenopus 1
29	18	7.4	2307	1	AF011388	AB012111 Clostridi
30	18	7.4	4161	4	AB014462	AB012112 Clostridi
31	18	7.4	4525	6	AB012111	AB022218 Arabidops
32	18	7.4	11584	6	AB012112	AL022605 Arabidops
33	18	7.4	12297	6	CBGBONT	AL035708 Arabidops
34	18	7.4	82893	7	AB022218	X70810 E.gracilis
35	18	7.4	88343	7	ATP19P19	D25323 Proso mille
36	18	7.4	114906	7	CLEGGCA	AB020877 Homo sapi
37	18	7.4	143172	7	PNTAATB	D87009 Human (lamb
38	18	7.4	9014	7	PNTAATB	AL031775 Human DNA
39	18	7.4	100000	9	AB020877	Z96074 Human DNA s
40	18	7.4	104386	9	D87009	AC002069 Human BAC
41	18	7.4	102200	9	HS30M3	X68757 H.sapiens K
42	18	7.4	155661	10	HS399M14	AC007560 Homo sapi
43	18	7.4	158537	10	HSAC002069	
44	18	7.4	1263	10	HSKX113A	
45	18	7.4	174369	35	AC007560	

## ALIGNMENTS

RESULT 1	AF127036	2826 bp	mrna	PRI	16-APR-1999
LOCUS	AF127036				
DEFINITION	Homo sapiens calcium-activated chloride channel protein 1 (CACCI)				
ACCESSION	AF127036				
NID	94385468				
VERSION	AF127036.1	GI:4585468			
KEYWORDS					
SOURCE					
ORGANISM					
human.					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE					
AUTHORS	Agnel,M., Vernet,T. and Culouscou,J.-M.				
TITLE	Cloning of three human homologs of bovine epithelial chloride				
JOURNAL	channel				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 2826)				
TITLE	Agnel,M. and Culouscou,J.-M.				
JOURNAL	Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des				
FEATURES	Carrières, Reuil-Malmaison 92500, France				
	Location/Qualifiers				

source  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="CACCI"  
5. .2749  
/gene="CACCI"  
/note="Dovine epithelial chloride channel homolog"  
/product="calcium-activated chloride channel protein 1"  
/codon\_start=1  
/protein\_id="AAD25487.1"  
/db\_xref="GI:4585469"  
/db\_xref="PID:94585469"  
/translation="MGPFKSVFLLILHLEGLSNSLIQLNNGYEGIVVAIDPNV  
EDETLIQIKDWTQASLYLEFATGRFYKRVALLIPETWKADYRPLEYTKNA  
DYLVAESTPGNDPEPTTEOMNGCKEGERIHLPETIGKLAIEYGPGRAPVEMHA  
LRMGVDEYNDEKFEYLSNGRIQAVCSAGITGVYKCGGSCYTRCTFNKVTGL  
YKGECEYVLOSROTETKASIMFAQHVDSIEVECTEONHNKRNKQNKCNLRSTWEVI  
RDEDEKRTTPTTQPPNPFTSLQIGRIYCLVLDKSGMATGNRLNRLNAGOLF  
LQTVELGSMVGVATFDSAAHVOSGLIQLNSGSDRDLAKRLPAAGSGTSGISGLRSL  
FTVIRKRYPTDGESEIVLTDGEDNTISGCFNEVKOSGAIHHTVALGPSAOLELSTK  
MTGGLQTVASDOVONNGILDAFGALSSGNAVSORSIOLESGLTLQNSQNMNGYIV  
DSVVGKDTLFLITWTTPQPIILMDPSGQKGSFYVDKTKMAYLQIPGIAKVTWKY  
SLQASSQTLTLVTSRASNAITLPITVTSKTKNDISKSPSPLVYANIRQASPLIRA  
SVALLESVNGKTVLELLDNGAGADATKDDGVYSRYFTYDNGRYSVKVRALGVN  
AARRVYIPQSGALYIPGMIENDEIOMNPPEIKNDVQHKQVCFSTSSGSEFVAS  
DVPNAPIDLPFGQITDLKAEIHGSSLINTWTAPGDYDHGTAHKYIIRISTSLD  
LRKFNEILOVNTTALIPKEANESEVLEFKPENITFENGTDLFLIAQAVDKYDKSEI  
SNIAVSLFIPQTPPETPSPDETSAPCPNHINSTIPGIIHLIKMKWIGELQSLA  
"

BASE COUNT 875 a 623 c 632 g 696 t

ORIGIN

Query Match 88.0%; Score 213; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 2.9e-100; Mismatches 0; Indels 0; Gaps 0;  
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTTATCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAAAGCTGTCTCCTTG 60  
DB 2596 GTTATCTCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAAAGCTGTCTCCTTG 2655

QY 61 TCCTATATTCATATCAACAGACACCTCCGCTGCAATTAATAAATATATGGA 120  
DB 2656 TCCTATATTCATATCAACAGACACCTCCGCTGCAATTAATAAATATATGGA 2715

QY 121 GTGATAGGAGAACTGCACTGTCAATAGCCTAGGCTGAATTTTGTGAGATAATAA 180  
DB 2716 GTGATAGGAGAACTGCACTGTCAATAGCCTAGGCTGAATTTTGTGAGATAATAA 2775

QY 181 ATAAATCATTCATCTTTTGTGATTAATAA 213  
DB 2776 ATAAATCATTCATCTTTTGTGATTAATAA 2808

RESULT 2  
AF039400 3311 bp mRNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,  
DEFINITION complete cds.  
ACCESSION AF039400  
NID 94009457  
VERSION AF039400.1 GI:4009457  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Eldle,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Paul,I.B.O.  
TITLE Genomic cloning, molecular characterization, and functional

analysis of human CLCA1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
Genomics 54 (2), 200-214 (1998)  
99047526  
2 (bases 1 to 3311)  
Gruber,A.D., Eldle,R. and Paul,I.B.O.  
Direct Submission  
Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/gene="hCLCA1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPFKSVFLLILHLEGLSNSLIQLNNGYEGIVVAIDPNV  
EDETLIQIKDWTQASLYLEFATGRFYKRVALLIPETWKADYRPLEYTKNA  
DYLVAESTPGNDPEPTTEOMNGCKEGERIHLPETIGKLAIEYGPGRAPVEMHA  
LRMGVDEYNDEKFEYLSNGRIQAVCSAGITGVYKCGGSCYTRCTFNKVTGL  
YKGECEYVLOSROTETKASIMFAQHVDSIEVECTEONHNKRNKQNKCNLRSTWEVI  
RDEDEKRTTPTTQPPNPFTSLQIGRIYCLVLDKSGMATGNRLNRLNAGOLF  
LQTVELGSMVGVATFDSAAHVOSGLIQLNSGSDRDLAKRLPAAGSGTSGISGLRSL  
FTVIRKRYPTDGESEIVLTDGEDNTISGCFNEVKOSGAIHHTVALGPSAOLELSTK  
MTGGLQTVASDOVONNGILDAFGALSSGNAVSORSIOLESGLTLQNSQNMNGYIV  
DSVVGKDTLFLITWTTPQPIILMDPSGQKGSFYVDKTKMAYLQIPGIAKVTWKY  
SLQASSQTLTLVTSRASNAITLPITVTSKTKNDISKSPSPLVYANIRQASPLIRA  
SVALLESVNGKTVLELLDNGAGADATKDDGVYSRYFTYDNGRYSVKVRALGVN  
AARRVYIPQSGALYIPGMIENDEIOMNPPEIKNDVQHKQVCFSTSSGSEFVAS  
DVPNAPIDLPFGQITDLKAEIHGSSLINTWTAPGDYDHGTAHKYIIRISTSLD  
LRKFNEILOVNTTALIPKEANESEVLEFKPENITFENGTDLFLIAQAVDKYDKSEI  
SNIAVSLFIPQTPPETPSPDETSAPCPNHINSTIPGIIHLIKMKWIGELQSLA  
"

BASE COUNT 1028 a 692 c 742 g 849 t

ORIGIN

Query Match 83.5%; Score 202; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 1.5e-94; Mismatches 0; Indels 0; Gaps 0;  
Matches 202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTTATCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAAAGCTGTCTCCTTG 60  
DB 2943 GTTATCTCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAAAGCTGTCTCCTTG 3002

QY 61 TCCTATATTCATATCAACAGACACCTCTGCAATTAATAAATATATGGA 120  
DB 3003 TCCTATATTCATATCAACAGACACCTCTGCAATTAATAAATATATGGA 3062

QY 121 GTGATAGGAGAACTGCACTGTCAATAGCCTAGGCTGAATTTTGTGAGATAATAA 180  
DB 3063 GTGATAGGAGAACTGCACTGTCAATAGCCTAGGCTGAATTTTGTGAGATAATAA 3122

QY 181 ATAAATCATTCATCTTTTGTGATTAATAA 202  
DB 3123 ATAAATCATTCATCTTTTGTGATTAATAA 3144

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
DEFINITION complete cds.

VERSION	KEYWORDS	SOURCE	ORGANISM
AF039401			Homo sapiens.
NID			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AF039401.1	GI:4009459		Euthelia; Primates; Catarrhini; Homnidae; Homo.
REFERENCE			1 (bases 1 to 35278)
AUTHORS			Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.
TITLE			Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca <sup>2+</sup> -activated Cl- channel proteins
JOURNAL			Genomics 54 (2), 200-214 (1998)
MEDLINE			2 (bases 1 to 35278)
REFERENCE			Gruber,A.D., Elble,R. and Pauli,B.U.
AUTHORS			Direct Submission
TITLE			Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
JOURNAL			Location/Qualifiers
FEATURES			1..35278
source			/organism="Homo sapiens"
			/db_xref="taxon:9606"
			/chromosome="1"
			/map="ip22-p31"
			1..1617
			/gene="hCLCA1"
			1..33522
			/gene="hCLCA1"
			767..1265
			/rpt_family="I1 transposable element"
			/rpt_type-dispersed
			1587..1596
			/gene="hCLCA1"
			/note="putative"
			1618..1708
			/gene="hCLCA1"
			/number=1
			joh(1618..1708,1962..2383,6667..6807,6979..7126,
			9696..9801,15454..15631,18591..18809,19783..20010,
			22253..22427,24522..24628,26644..26859,27447..27708,
			28766..28936,31840..32079,32919..33522)
			/gene="hCLCA1"
			/product="calcium-dependent chloride channel-1"
			1962..2383
			/gene="hCLCA1"
			/number=2
			joh(12222..2383,6667..6807,6979..7126,9696..9801,
			15454..15631,18591..18809,19783..20010,22253..22427,
			24522..24628,26644..26859,27447..27708,28766..28936,
			31840..32079,32919..33310)
			/gene="hCLCA1"
			/codon_start=1
			/product="calcium-dependent chloride channel-1"
			/protein_id="AAC95429.1"
			/db_xref="PID:94009460"
			/db_xref="GI:4009460"
			/translation="MEPFKRSVFLILHLLEGALNSLQLANNNGYICIVADPNP
			EEDELIIQDMVTQASLYLEATGKRFFENNAVILLPTWTKRADYVRRLLEYRNA
			LVAEASTPGNDEPYTEOMGCNGEKERIHLPDFTAGRKLEYGCGAFAFEMAH
			LRMVFPEYNDEKEFYLSNGRIQAVRCSAGITGVNVKKCGSGCYTRCTFNVTSL
			YEKCSEFILOSROFERKASIMEPAOHVDSTVECPTEONNRKAPNQNKCRSTWEVI
			RDSDEFRTKTPMTQTNPPTSILQIQORICYVLDSGSMATGNRNLRNLAQGFL
			LQYELSSWGVTFDSSAAHVOSELIIQINSSDDTLAKRLPPAASGGTISCSLRSA
			FTVIRKRPDGESEIVLLTGEDNTIGSCFEVNOQSAILHTVALGSSAOOELESK
			MTGLQYASDOVONNLIDAFGALSNGNVSORSILSKSGLTLLNSGMNGETVIV
			DSTVGKDLFTLTWTQPOLIPMDPSGOKGGVVNKNRMALQIPGAKGTWXY
			SIQASNOTLITVSRASNALPIPIITYTSKTNDTSKEPSPVYVYANIGAQPILLIA
			SVTLAISVNKTYTTELINDGAQADTKDDGYSTRFTTYTDNGRSVYVRALGCVN
			AARRVVIPDGSGALYPGWIMENDIQNNPPPELNKDQVDVKQVCFSRTSSGGSFAS
			DVPARRVLPDLFGQITDNLNIEHGSSINTLTWAPDPDDVGHGANHYIIRISTSID

exon	6667..6807	/gene="hclcal"	
exon	6979..7126	/number=3	
exon	9696..9801	/number=4	
exon	9896..9901	/gene="hclcal"	
exon	15454..15631	/number=5	
exon	18591..18809	/gene="hclcal"	
exon	19783..20010	/number=7	
exon	22253..22427	/number=8	
exon	24522..24628	/gene="hclcal"	
exon	26644..26859	/number=10	
exon	27447..27708	/number=11	
exon	28766..28936	/gene="hclcal"	
exon	31840..32079	/number=13	
exon	32919..33522	/gene="hclcal"	
exon	33215	/number=15	
BASE COUNT	11245 a 6870 c 7230 g 9933 t		
ORIGIN			
Query Match	66.9%; Score 162; DB 11; Length 35278;		
Best Local Similarity	100.0%; Pred. No. 7.6e-74;		
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY 41 GATGAAGCTGCTCTCTGCTCTTAATATCATATCAACAGCACCATTCTGGCATTTCAC 100			
DB 33197 GATGAAGCTGCTGCTCTCTGCTCTTAATATCATATCAACAGCACCATTCTGGCATTTCAC 33256			
QY 101 ATTTTAAAAATTANGTGAAGTGCATGAGACACTGCAGCTGTCATATAGCCTAGGGCTGA 160			
DB 33257 ATTTTAAAAATTATGTGAAGTGCATGAGACACTGCAGCTGTCATATAGCCTAGGGCTGA 33316			
QY 161 ATTTTGTGAGTAAATAATAATATCATCTTCATCTTTTTTT 202			
DB 33317 ATTTTGTGAGTAAATAATAATATCATCTTCATCTTTTTTT 33358			
RESULT 4			
LOCUS 195746			
DEFINITION Sequence 8 from patent US 5733748.			
ACCESION 195746			
NID 93940216			
VERSION 195746.1			
KEYWORDS GI:3940216			
SOURCE Unknown.			
ORGANISM Unknown.			



from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://genomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRail are annotated as misc features.

```
FEATURES
Source
1. 94038
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="11"
/map="C1C06C07"
/clone="F21P24"
1. 22544
/note="overlap with BAC clone T20K9
(AC004786:64143..86686)."
<1..>1327
/gene="F21P24.1"
/note="T20K9.16"
join(1..92,168..326,400..571,679..977,1088..1327)
/gene="F21P24.1"
/codon_start=3
/product="putative Ca2+-ATPase, 5' partial"
/protein_id="AACI1813.1"
/db_xref="PID:93169170"
/db_xref="GI:3169170"
/translation="MTDDGAIIEGPFVREKNOEMLEIPKIOVMARSSPMKRTLY
KQRTPEVAVVAGDGTNDAPALHEADIGLACGAEVPAEADIVITIDNESTIV
TAAWGRSVYINIQFVQLTIVNVALIVNSACLTGSAFLIVQILWAMIDTL
GALATKRPNNELMRKMPVGRGNFTNAMRNILGOAVYQFIITWILQAGKSMFG
LVGDSITLVNTLINCFCVQGVNESSREMEIDVKGIDNVTFFVVGATVFEQ
IILIEFGTFASTPLTIYQVFISFVGLGMPVIAAGLKKIPV"
join(<1..92,168..326,400..571,679..977,1088..>1327)
/gene="F21P24.1"
MRNA
1903..2036
/repeat_region
/misc_feature
/complement(2191..2239)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
misc_feature
/complement(2295..2360)
/note="exon predicted by xgrail, quality good_shadowexon"
2455..4493
/gene="F21P24.2"
/note="T20K9.17; similar to serine carboxypeptidase I"
/pseudo
repeat_region
/complement(3328..3385)
/rpt_family="AT_rich"
repeat_region
/complement(3390..3506)
/rpt_family="TAAAAn"
/misc_feature
4509..4538
/note="exon predicted by xgrail, quality marginal"
repeat_region
/complement(4553..4590)
/rpt_family="(TA)n"
/rpt_family="(TA)n"
repeat_region
/complement(4633..4653)
/rpt_family="AT_rich"
/misc_feature
/complement(4830..4954)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
4848..4958
/misc_feature
```

```
/note="exon predicted by xgrail, quality excellent"
complement(5440..5613)
/note="exon predicted by xgrail, quality good_shadowexon"
101n(6436..6560,6635..6780,6881..6957,7038..7160,
7565..7667,7736..7773,7979..8077,8179..8265,8358..8426,
8612..8683,8764..8885,9346..9459,9572..9617,9705..>9785)
/gene="F21P24.3"
<6436..>9785
/gene="F21P24.3"
/note="T20K9.18"
101n(6436..6560,6635..6780,6881..6957,7038..7160,
7565..7667,7736..7773,7979..8077,8179..8265,8358..8426,
8612..8683,8764..8885,9346..9459,9572..9617,9705..>9785)
/gene="F21P24.3"
/codon_start=1
/product="putative serine carboxypeptidase I"
/protein_id="AACI1814.1"
/db_xref="PID:93169171"
/db_xref="GI:3169171"
/translation="MELTKLVLLFLINHHVSGSIVKLPFEGPLPELENGYI
GIEEEVQLFYYFIKSRNPREDLLMLSGGPCSSITGLFNGPLAKSEVYNG
SVPSLAVSTYWTAKTANITFLDQVAGFSYRAPLIDPTDGTGKRIHFLQKML
KHPOSSNHFYAGDSYSGMIVPALVOEISKGNVICNPILKGVVGNITHEDDP
NRIPEHSGMALISDELVESTIREACKGVFNPDPNTKCIKVEEFHCTOKLNEFI
LSPDCDTPSPCYLTPFLYLSFPMANDESRALAHYKRSIGMEECNYSKPYNDIK
SSVPTMNNVSQVRSILYSGDHDLVFPLPILQAIKLNYSIIDEMRPWMIRDQITG
YTRYSNKMTFAYVKSQHTAEKQESIFMERWINGOPL"
complement(7277..7399)
/rpt_family="AT_rich"
9116..9138
/rpt_family="AT_rich"
10479..10527
/note="exon predicted by xgrail, quality good"
101n(<10590..10717,10800..10945,11012..11088,11168..11290,
11602..11704,11806..11828,11912..12013,12378..12464,
12552..12620,12727..12798,13609..13730,14190..14303,
14407..14452,14559..>14639)
/gene="F21P24.4"
<10590..>14639
/note="F21P24.4"
101n(10590..10717,10800..10945,11012..11088,11168..11290,
11602..11704,11806..11828,11912..12013,12378..12464,
12552..12620,12727..12798,13609..13730,14190..14303,
14407..14452,14559..14639)
/gene="F21P24.4"
/codon_start=1
/product="putative serine carboxypeptidase I"
/protein_id="AACI1815.1"
/db_xref="PID:93169172"
/db_xref="GI:3169172"
/translation="MSLTFELLLIVILSHAHSGSIVFLPGEPLPELEETGY
IGIEEEVQLFYYFIKSRNPREDLLMLSGGPCSSITGLFNGPLAKSEVYNG
SVPSLAVSTYWTAKTANITFLDQVAGFSYRAPLIDPTDGTGKRIHFLQKML
SKHQGFSPNFTYVGDSTSGMIVPALVOEISKGNVICNPILKGVVGNITHEDDP
LIPYAGMALISDELKSMERCKGNVAVDSLNKCYLIDYQCIHKLNNYHILPL
DODITSDFCFLRYTLITFMANNKSVRALOVNKSIGKRNKNSYNNDIKSV
AYHKNSIDGYSRLIYNGDHDMMVYFLTAQVIRLANSITDDMPFMMIINOIAGYR
SYSNKMTFATIKSGHTAEVKEKESIMFKWISNOPL"
13918..14008
/rpt_family="(TA)n"
/rpt_family="(TA)n"
repeat_region
/complement(14826..14899)
/note="exon predicted by xgrail, quality good"
complement(17227..17330)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
17429..17488
/note="exon predicted by xgrail, quality marginal"
101n(<17954..18072,18144..18289,18426..18502,18632..18751,
19006..19108,19205..19242,19338..19439,19871..19957,
20049..20126,20326..20397,20476..20594,21039..21152,
21262..21307,21406..>21486)
/gene="F21P24.5"
```





```

misc_feature /rpt_family="AT_rich"
               complement(5810..5841)
               /note="exon predicted by xgrall, quality excellent"
repeat_region complement(6101..6133)
               /rpt_family="AT_rich"
               6962..6993
repeat_region /rpt_family="AT_rich"
               7017..7101
misc_feature /note="exon predicted by xgrall, quality excellent"
               7426..7502
misc_feature /note="exon predicted by xgrall, quality marginal"
               complement(7587..7695)
               /note="exon predicted by xgrall, quality
               marginal_shadowexon"
misc_feature 8000..8041
               /note="exon predicted by xgrall, quality excellent"
               8240..8272
repeat_region /rpt_family="AT_rich"
               8319..8431
               /rpt_family="(TA)n"
misc_feature 8514..8586
               /note="exon predicted by xgrall, quality excellent"
               complement(8517..8654)
misc_feature /note="exon predicted by xgrall, quality
               excellent_shadowexon"
               8602..8775
misc_feature /note="exon predicted by xgrall, quality good"
               9289..9453
               /note="exon predicted by xgrall, quality marginal"
               complement(9516..9678)
misc_feature /note="exon predicted by xgrall, quality good_shadowexon"
               9759..9946
               /note="exon predicted by xgrall, quality good"
               complement(9764..9868)
misc_feature /note="exon predicted by xgrall, quality good_shadowexon"
               complement(10009..10083)
               /note="exon predicted by xgrall, quality
               marginal_shadowexon"
repeat_region 10522..10608
               /rpt_family="AT_rich"
               10794..10894
               /note="exon predicted by xgrall, quality excellent"
               complement(11119..11164)
repeat_region /rpt_family="AT_rich"
               11379..11486
               /rpt_family="(TA)n"
               11501..11669
               /note="exon predicted by xgrall, quality excellent"
               complement(11880..11902)
repeat_region /rpt_family="AT_rich"
               11966..12017
               /rpt_family="(CAA)n"
               12561..12672
               /note="exon predicted by xgrall, quality good_shadowexon"
               complement(12694..12722)
               /rpt_family="AT_rich"
               12974..13067
               /note="exon predicted by xgrall, quality
               marginal_shadowexon"
               complement(join<13494..14478,14567..14727,14795..14926,
               15029..15175)
               /gene="T20K9.2"
               complement<13494..>15175)
               /gene="T20K9.2"
               /note="T30L20.7: identical to GB:U23481"
               complement(join(13494..14478,14567..14727,14795..14926,
               15029..15175))
               /gene="T20K9.2"
               /codon_start=1
               /product="1-aminocyclopropane-1-carboxylate synthase
               (ACS4)"
               /protein_id="AAC32428.1"

```

```

Query Match      8 7%: Score 21; DB 8; Length 86686;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 183 AATCATTCATCCCTTTT 203
    |||||
DB 73683 AATCATTCATCCCTTTT 73703

```

```

RESULT 8
LOCUS CEY19D2 179854 bp DNA HTG 14-OCT-1998
DEFINITION Caenorhabditis elegans chromosome II clone Y19D2, WORKING DRAFT
SEQUENCE, in unordered pieces.
ACCESSION AL021347
NID 93646815
AL021347.1 GI:3646815
KEYWORDS HTG: HTGS, PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 179854)
AUTHORS White, S.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1HQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jesse.sanger.ac.uk or rvenematode.wustl.edu
On Sep 24, 1998 this sequence version replaced gi:2804161.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

FEATURES
    source          location/Qualifiers
                    1..179854
                    /organism="Caenorhabditis elegans"
                    /db_xref="taxon:6239"
                    /chromosome="II"
                    /clone="Y19D2"

```

```

BASE COUNT 56786 a 32609 c 32628 g 56230 t 1601 others
ORIGIN

```

```

Query Match      8 7%: Score 21; DB 34; Length 179854;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 211 AATTTCTAAATGATTTT 231
    |||||
DB 74101 AATTTCTAAATGATTTT 74081

```

```

RESULT 9
LOCUS CEM106 39973 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid M106, complete sequence.
ACCESSION Z46935
NID 9602783
Z46935.1 GI:602783
KEYWORDS HTG; conglutin; F-actin capping protein; GMP synthase; nucleotide
binding; guanine oxidoreductase; transfer RNA; tRNA-Leu.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

```

REFERENCE	1 (bases 1 to 39973)
AUTHORS	Palmer,S.
TITLE	Direct Submission
JOURNAL	Submitted (16-DEC-1994) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or tw@nemalode.wustl.edu
REFERENCE	2 (bases 1 to 39973)
AUTHORS	Wilson,R., Alnsoough,R., Anderson,K., Baynes,C., Berke,M., Bofield,J., Burton,M., Dear,S., Du,Z., Durkin,R., Favallo,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,N., Jier,M., Johnston,L., Jones,M., Kersey,J., Kirsten,O., Laister,N., Lafreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopa,A., Saunders,D., Showkeen,R., Smaldon,N., Smith,A., Sonhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Woldman,P.
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL	Nature 368 (6466), 32-38 (1994)
MEDLINE	94150718
COMMENT	Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see:- <a href="http://webace.sanger.ac.uk/cgi-bin/display?db=wormacc&amp;class=Sequence&amp;object=M106">http://webace.sanger.ac.uk/cgi-</a> <a href="http://webace.sanger.ac.uk/cgi-bin/display?db=wormacc&amp;class=Sequence&amp;object=M106">bin/display?db=wormacc&amp;class=Sequence&amp;object=M106</a> Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring subclones. IMPORTANT: This sequence is not the entire insert of clone M106. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring subclisions. The true left end of clone M106 is at 1 in this sequence. The true right end of clone M106 is at 2020 in sequence Z36949. The true left end of clone F59E10 is at 39873 in this sequence. The start of this sequence (1..105) overlaps with the end of sequence A1031266. The end of this sequence (39873..39973) overlaps with the start of sequence Z36949.
FEATURES	Location/Qualifiers
source	1..39973
gene	/organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="II" /clone="M106" complement(110154..16153) /gene="M106.2" complement(join(10154..10357,10904..11026,11275..11466, 12100..12178,13101..13253,13649..13766,13809..13946, 14493..14716,14781..14970,16014..16153)) /gene="M106.2" /note="weak similarity to conglutin (PIR accession number A33090): CDNA EST EMBL:F01355 comes from this gene; CDNA EST YK383f3.3 comes from this gene" /codon_start=1 /protein_id="CAA87049.1" /db_xref="PID:e151703" /db_xref="PID:g3878712" /db_xref="GI:3878712" /db_xref="SPTR:EML:009592" /translation="MDGNPICRNLTLSVTHAKNCRKNSLKLCSTIRACLEIIM DMELDIIVLEIKSTISLANFNQATSISSGTSHTSOOSTDSVDLNTIFRSV NQSDTPELIASSFESEKDMEFTPTETIAINFTSGTPVTIPPSYTDADNLIDTKAT"

gene  
CDS  
complement(24287..29464)  
/gene="M106.3"  
complement{join(24287..24394,24693..24749,24799..24885,  
25598..25732,25820..25908,26455..26545,27093..27242,  
27299..27439,27941..28087,28139..28363,29365..29464)}  
/note="weak similarity with quinone oxidoreductase; cDNA  
EST EMBL:C13104 comes from this gene; cDNA EST yk264f3.5  
comes from this gene"  
/codon\_start=1  
/protein\_id="CAAB87050.1"  
/db\_xref="PID:e1351704"  
/db\_xref="PID:g3878713"  
/db\_xref="GI:3878713"  
/translation="MRIFQAROMTSSSSAHOPLPADNKVVFIVKRGGENAPITDFC  
DYPIDGASCTLEVDLYTGIRLTMTWKTRFIVCSDDSLKYNNLSRRNSFSVLISG  
VGVVERVSAGCTLEVDLYTGIRLTMTWKTRFIVCSDDSLKYNNLSRRNSFSVLISG  
ALSGITALLGRKKALIDRSRPTIIVSGAOSGSGLAQIARLEGSCRVIGSD  
DCYQLIKREFGNVDINPKTEWVSRLGAEGEIDYNDVNGVJSDVIYANNNEG  
RYVLGGQIAVNTDLPPPELPETHTKIIKENNIORELVDMYDEIDEAVAANSEW  
LOODKIKEYETIYDGILNAPSFAVDMNMKNKGMLIHSHSKENLFNFKEHREEF  
GNSVKYEIMELVKTRETMLTNOEVLSFNALKEDDNQOHEKROCKSTVIEFKY  
LMTSPAVNVSECELPPEPKR"  
complement(32748..37364)  
/gene="M106.4"  
complement{join(32748..32852,32904..33452,33507..34096,  
34145..34504,34647..35159,35284..35344,36894..36998,  
37199..37261,37332..37364)}  
/gene="M106.4"  
/note="similar to GMP synthase; cDNA EST EMBL:M68983 comes  
from this gene; cDNA EST EMBL:D27938 comes from this gene;  
cDNA EST EMBL:D27937 comes from this gene; cDNA EST  
EMBL:D37506 comes from this gene; cDNA EST yk200f6.5 comes  
from this gene; cDNA EST yk208d1.5 comes from this gene;  
cDNA EST yk356d1.5 comes from this gene; cDNA EST  
yk486f1.5 comes from this gene"  
/codon\_start=1  
/protein\_id="CAAB7052.1"  
/db\_xref="PID:e1351706"  
/db\_xref="PID:g3878715"  
/db\_xref="GI:3878715"  
/translation="MKFSOYPCPSRKPRKVGICARGSKRHCRULPSELQARKTRAVSK  
VSEHTMKSSSLMDINESOHSTNKAPPKPAEDRFPSANNNGSHVLVELPEKV  
SSGEAILIDLEFAQGKVIDRWRELVLQOSEMFNPNTARTLIELGGESVILISGPN  
SVFEEDAPSIDPEFTCGIPYLVCYGOLMMKNGCVGTREHIIEDEACELQDVTSV  
HLFNLHTEETVELTHGDVSSEAYAPAPRTKVAKSGHNAGJCNRNLRVIGQVDFE  
DTTGTGTAEMEVELEFKVWGCGNETIOWREDCSEIYNVIGDKKLVMVSGSNVAV  
CAALIRALGPBRTVALHINDGFMRHSDKAVERSIALADPIHYNGCTPFRESSEH  
CKDSVALDECDDPEMKRKTIIGNTFIRKDYIMKDLDINHBEYFLAQSTLRBDLIAEA  
SALAGHADITKEHNDITFVLEKRLKAVEPDKDFKDEVEELKGLGPESTVRA  
HPFPDQAIRILCASDRPELMFEMPYNDVOGRITDIVYEOFANRSAINSGIYEH  
LOFERARNNVLTISEDEROLAQOOTFOISHAIPIKTVGOGARSXVAALSTEQ  
RIPIPKLFEAAVASVPKLEHGVRNVYAPVGLGIESIEDLTRTYVPHITVLIKQADH  
IANTLIERVAASGTOLPNGHKRQQVADPVLLPDPRDRDMIGSYRHSIYLRPVT  
SDFMGQAATGYVHPEITLIEMDKAITNWLGISRYLLDMTCACPITTEME"  
3787)..3891"  
/gene="cap-2"  
join(37878..37789,37838..37955,38005..38098,38147..38260,  
38313..38463,38516..38620,38666..38791,38870..38971)  
/gene="cap-2"  
/note="similar to F-actin capping protein beta subunit;  
cDNA EST EMBL:T00119 comes from this gene; cDNA EST  
EMBL:T00118 comes from this gene; cDNA EST yk325a2.5 comes  
from this gene; cDNA EST yk456a6.5 comes from this gene;

cdna EST yk471e9.5 comes from this gene"

/codon\_start=1  
/protein\_id="CA87051.1"  
/db\_xref="PID:el351705"  
/db\_xref="PID:g387174"  
/db\_xref="GI:3878714"  
/db\_xref="GI:3878714"  
/db\_xref="SWISS-PROT:P34686"  
/translation="MGEOQLDCAIDLMKRLRPQCDKRLTDLICLPHLVNDLSTID  
QPLKIAADRETGQYLLCDYNRDGDSRSPSNNTYDPLEEGQLPSEKRRMELEANA  
AFESYRDLYEGGVSVYFMDLDNGFAGIYLIRKEGQDAPINDQNDLHANGMIED  
ARQAHYKLTSTIMLMQNTKSSGVMMIGSLTIRKEHEDAPINDQNDLHANGMIED  
QESKRLTINITYGKTKKVSDDLSTSEKQSELEKQDIRELNNAANRN"  
join(complement(236949.1:1212.1290),  
complement(236949.1:1036.1156),  
complement(236949.1:203.389),  
complement(236949.1:102.146),complement(39640.39973),  
complement(39385.39590))  
/gene="cyp-4"  
/note="similar to cyclophilin; cdna EST yk464d7.5 comes  
from this gene"

/codon\_start=1  
/protein\_id="CA87053.1"  
/db\_xref="PID:el351707"  
/db\_xref="PID:g387176"  
/db\_xref="GI:3878716"  
/db\_xref="GI:3878716"  
/translation="MGKQKQKQKLYLTSTWKSIGCHKDTRQLQRAQKRLPINH  
CSLSLPEDPVCARSGEIEFDLTAIVPLKHKHNPCTGKPLVAKDLHLKFDGEGD  
KFCQPVTFRTDHSILALITSGVNSHEVQELNRLKDLTDVPTRADLID  
LDQNHLEKFMEOPLHVKDLTKTSEEKEDAMKDEKFIIRNNACKSVLDLID  
EYPRKSTETDEADEINAHYSGQKVAAGTSVAPVSNKAAYLNDNTVRSRY  
KKAFVRLYTFEGPLNELFAPKPKACENTITCSGYNNTFPHLIRKPMIOGD  
PTGTHGSGESIMDKPFSDEFISGSHDKRGVLSANKSNTNGSOFTTFPCYTLDR  
KHITFGRLVGGDTLTLEKLETEGIDVPVSVIMAEFVDPFEAEAEVQAEERA  
ELIKTNRDAISLANKKAKETATKPEAVGTGVGYMSSAAVNRQGMEDVPLEAAK  
KTFERAGLIGDFSKW"  
join(246794.1:39146.39239,246794.1:39326.39409,  
246794.1:39569.40500,246794.1:40552.40703,  
246794.1:41164.41269,246794.1:41319.41452,  
AL031266.1:107.112,AL031266.1:912.1068,453.614,  
5546.5638,5736.6328,6853.7102,8138.8410,8672.8797,  
9022.9321,9504.9776)  
/gene="M106.1"  
/note="similar to chromosome segregation protein; cdna EST  
y106.1"

Query Match 8.7%; Score 21; DB 36; Length 39973;  
Best Local Similarity 100.0%; Pred. No. 0.8;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 211 AAATTTCTAAATGATATTTT 231  
|||||  
Db 27521 AAATTTCTAAATGATATTTT 27501

RESULT 10  
AB025615/c 83544 bp DNA PLN 07-APR-1999  
LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone: K5K13,  
DEFINITION complete sequence.  
ACCESSION AB025615  
NID 94589421  
VERSION AB025615.1 GI:4589421  
KEYWORDS HTG.  
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:1ld:Mitsui PI  
clone:K5K13.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
Arabidopsis;  
1 (sites)  
REFERENCE Nakamura,Y.  
AUTHORS Structural Analysis of Arabidopsis thaliana Chromosome 3. II  
JOURNAL Unpublished (1999)

REFERENCE 2 (bases 1 to 83544)  
AUTHORS Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-APR-1999) to the DDBJ/EMBL/Genbank databases.  
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of  
Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan  
(E-mail:nakamukazusa.or.jp, Tel:+81-438-52-3935,  
Fax:+81-438-52-3934)  
FEATURES  
source  
1. 83544  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="3"  
/clone="K5K13"  
/clone\_1ld="Mitsui PI"  
BASE COUNT 28677 a 14395 c 13562 g 26910 t  
ORIGIN

Query Match 8.3%; Score 20; DB 7; Length 83544;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 213 AATTTCTAAATGATATTTT 232  
|||||  
Db 82992 AATTTCTAAATGATATTTT 82973

RESULT 11  
AC002994/c 143819 bp DNA PRI 10-NOV-1997  
LOCUS Homo sapiens chromosome 17, clone HRPc987K16, complete sequence.  
DEFINITION AC002994  
ACCESSION G2599233  
NID AC002994.1 GI:2599233  
VERSION HTG.  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 143819)  
REFERENCE Birren,B., Fasmann,K., McKernan,K., Munro,C., Nusbaum,C.,  
Richardson,P., and Lander,E.  
Homo sapiens chromosome 17, clone HRPc987K16  
Unpublished  
2 (bases 1 to 143819)  
Hawkins,T.L., Birren,B.W., Fasmann,K.H., Nusbaum,C., Lander,E.S.,  
McKernan,K., Munro,C., Richardson,P., Baldwin,J., Barna,N.,  
Cantu,C., Chang,A., Cooke,P., Daly,M.J., Devon,K., Dewar,K.,  
Durette,B., Forrest,C., Gage,D., Gensheimer,S., Gerald,K.,  
Gilmartin,T., Hagos,B., Halphen,I., Harris,K., Howland,J.C.,  
Huang,J., Hui,L., Jacotot,L., Kirby,A., Lane,M., Mackenzie,J.,  
Margolis,N., McDermott,J., Molla,M., Morrow,J., Nachman,A.,  
Naylor,J., Nusbaum,C., O'Connor,T., Oltow,A., Peterson,K.,  
Reeve,M.P., Roberts,D., Rollins,G., Sarnak,A., Shiu,P., Shyam,R.,  
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,  
Wilmer,F., Zemtseva,I. and Zody,M.  
Direct Submission  
TITLE Submitted (06-OCT-1997) Whitehead Institute/MIT Center for Genome  
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 143819)  
REFERENCE Birren,B., Fasmann,K., McKernan,K., Munro,C., Nusbaum,C.,  
Richardson,P., Lander,E., Baldwin,J., Barna,N., Cantu,C., Chang,A.,  
Cooke,P., Daly,M.J., Devon,K., Dewar,K., Durette,B., Forrest,C.,  
Gage,D., Gensheimer,S., Gerald,K., Gilmartin,T., Hagos,B.,  
Halphen,I., Harris,K., Howland,J.C., Huang,J., Hui,L., Jacotot,L.,  
Kirby,A., Lane,M., Mackenzie,J., Margolis,N., McDermott,J.,  
Molla,M., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Oltow,A.,  
Peterson,K., Roberts,D., Rollins,G., Sarnak,A., Shiu,P., Shyam,R.,  
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,  
Zemtseva,I. and Zody,M.  
Direct Submission

## JOURNAL

Submitted (10-NOV-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Nov 10, 1997 this sequence version replaced g1:2538080.  
 The Staden databases, finishing information, and all chromatographic files used in the assembly of this clone are available from our anonymous ftp site.

## COMMENT

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.  
 Location/Qualifiers

## FEATURES

## source

```

1. 143819
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="Peter de Jong/ human PAC library"
/map="17"
/chromosome="17"
repeat_region complement(404..452)
/rpt_family="CAAAA)n"
repeat_region complement(455..706)
/rpt_family="AluSx"
repeat_region complement(1273..1348)
/rpt_family="AT_rich"
repeat_region 1440..1472
/rpt_family="AT_rich"
repeat_region 2856..2945
/rpt_family="AT_rich"
repeat_region 2956..2995
/rpt_family="CAAAA)n"
repeat_region 2996..3094
/rpt_family="MER58B"
repeat_region 3093..3314
/rpt_family="MER44A"
repeat_region 3339..3385
/rpt_family="MER44B"
repeat_region 3420..3658
/rpt_family="MER58B"
repeat_region 3659..3748
/rpt_family="AluSx"
repeat_region 4460..4484
/rpt_family="MER58B"
repeat_region 5284..5559
/rpt_family="AT_rich"
repeat_region 5573..6010
/rpt_family="L1"
repeat_region 6024..6576
/rpt_family="L1"
repeat_region 6550..7043
/rpt_family="L1PA16"
repeat_region 7025..7269
/rpt_family="L1"
repeat_region 7292..7647
/rpt_family="L1"
repeat_region 7644..8145
/rpt_family="L1"
repeat_region 8147..8445
/rpt_family="L1PA15"
repeat_region 8466..9239
/rpt_family="AluSx"
repeat_region 9240..9265
/rpt_family="L1PA15"
repeat_region 9662..10379
/rpt_family="(CAA)n"
repeat_region 10482..11290
/rpt_family="L1M4"
repeat_region 11348..11888
/rpt_family="L1M4"
repeat_region 12043..12502
/rpt_family="MER68A"
repeat_region 12512..12796
/rpt_family="L1MB7"
repeat_region 12797..13383
/rpt_family="AluDb"

```

```

repeat_region 13403..13695
/rpt_family="L1MB7"
repeat_region 13701..14132
/rpt_family="AluSx"
repeat_region complement(14152..14433)
/rpt_family="L1MB7"
repeat_region complement(14505..14804)
/rpt_family="AluSg"
repeat_region complement(14805..14961)
/rpt_family="AluSg/X"
repeat_region 14962..17391
/rpt_family="L1MB7"
repeat_region 17444..17660
/rpt_family="L1"
repeat_region complement(18127..19262)
/rpt_family="L1MB1"
repeat_region complement(19263..19388)
/rpt_family="MER58B"
repeat_region complement(19384..19511)
/rpt_family="MER58B"
repeat_region complement(19528..19747)
/rpt_family="L1"
repeat_region 19756..20043
/rpt_family="AluSg"
repeat_region complement(20044..20226)
/rpt_family="L1"
repeat_region complement(20269..20580)
/rpt_family="L1PA5"
repeat_region complement(20571..20905)
/rpt_family="L1"
repeat_region complement(21290..21487)
/rpt_family="L1MB3/4"
repeat_region complement(21489..21779)
/rpt_family="AluDb"
repeat_region 22187..22464
/rpt_family="AluSx"
repeat_region 22466..22519
/rpt_family="(TAAA)n"
repeat_region complement(22524..22894)
/rpt_family="L1MB3/4"
repeat_region complement(23015..23315)
/rpt_family="AluDb"
repeat_region 23880..23930
/rpt_family="AT_rich"
repeat_region complement(24251..24432)
/rpt_family="AluDb"
repeat_region 24508..24553
/rpt_family="AT_rich"
repeat_region 24750..24794
/rpt_family="AT_rich"
repeat_region 25881..25939
/rpt_family="AT_rich"
repeat_region 26160..26208
/rpt_family="MIR"
repeat_region complement(26220..26455)
/rpt_family="MER58A"
repeat_region 26500..26750
/rpt_family="AluSg"
repeat_region 26755..26856
/rpt_family="AT_rich"
repeat_region 26892..26981
/rpt_family="MIR"
repeat_region complement(28359..28657)
/rpt_family="AluSg"
repeat_region 30006..30096
/rpt_family="MIR"
repeat_region 30278..30635
/rpt_family="THE1C"
repeat_region 30730..31050
/rpt_family="AluY"
repeat_region complement(31142..31447)
/rpt_family="AluSx"

```

repeat\_region 31723 .31781  
/rpt\_family="MIR"  
complement(33281..33320)  
repeat\_region /rpt\_family="AT-rich"  
complement(33477..33531)  
repeat\_region /rpt\_family="(GAAA)n"  
complement(33543..34783)  
repeat\_region /rpt\_family="LIP2"  
complement(34803..38450)  
repeat\_region /rpt\_family="L1"  
complement(38742..38872)  
repeat\_region /rpt\_family="FLAM\_A"

Query Match 8.3%; Score 20; DB 11; Length 143819;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 TTTTAAATATGATGAG 121  
Db 107984 TTTTAAATATGATGAG 107965

RESULT 12  
AE001386/c 12940 bp DNA INV 06-NOV-1998  
LOCUS Plasmodium falciparum chromosome 2, section 23 of 73 of the  
DEFINITION complete sequence.  
ACCESSION AE001386 AE001362  
NID 93845148  
VERSION AE001386.1 GI:3845148  
KEYWORDS  
SOURCE malaria parasite P. falciparum.  
ORGANISM Plasmodium falciparum  
REFERENCE 1 (bases 1 to 12940)  
AUTHORS Gardner, M.J., Tetteh, H., Carucci, D.J., Cummings, L.M., Arayand, L., Koolin, E.V., Shallow, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jung, J., Aston, C., Lal, Z., Schwartz, D.C., Petrea, M., Salzbeg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O., Fraser, C.M., Adams, M.D., Venter, J.C. and Hoffman, S.L.  
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum  
JOURNAL Science 282 (5391), 1126-1132 (1998)  
MEDLINE 99021743  
REMARK Eritrum: [[published erratum appears in Science 1998 Dec 4; 282(5395):1827]]  
REFERENCE 2 (bases 1 to 12940)  
AUTHORS Gardner, M.J.  
TITLE Direct Submission  
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA  
FEATURES  
source  
1..12940  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="2"  
complement(837..1791)  
/gene="PFB0305c"  
complement(join(837..1141,1278..1791))  
/gene="PFB0305c"  
/note="identified by sequence similarity"  
/codon\_start=-1  
/product="merozoite surface antigen MSP-5 (EGF domain)"  
/protein\_id="AAC71850.1"  
/db\_xref="PID:3845149"  
/db\_xref="GI:3845149"  
/translation="MTITLITLIVFVYFYSNLNKNKNTFLVRLNDRKGGEG  
TSKKNKNNNNNNNEKKEGSPKMEKNSDSDKPNDSHESKNSNSNT  
QKEPERKNSNRLDSSESSSATRSVDISEHNSNPPETKEGSEPPDLINENAE  
IQEPPNRLHFNVDDEVPHYSALRYNKEKNTDEMLLYNNMSDQNRSCAINNGC  
SDDOICININNGVCKICDKGYLGTGKCIILNSYSCHPFFSILIVITTLILFV"  
complement(2791..3753)  
/gene="PFB0310c"

CDS  
complement(join(2791..3125,3270..3753))  
/gene="PFB0310c"  
/note="identified by sequence similarity"  
/codon\_start=-1  
/product="merozoite surface antigen MSP-4 (EGF domain)"  
/protein\_id="AAC71851.1"  
/db\_xref="PID:3845150"  
/db\_xref="GI:3845150"  
/translation="MTIVFLIVHFFIICITINFDKLYISGSYNIVPENGRLNRL  
GEKPNVDVSTNTPGNESSASPNSDAEKDEKASQEGSEKKNESQSAN  
GSDVKEEKTNEKDDGKTDRKQVLEKSPKESQMDKKTETAIIPKVVQPS  
SGGVYGEEDHNEGEGHEHEHEDDDDDDDDDDDDDDDDDDDDDDDDDDD  
LCEYGNRRVCKCKEGYKLBCEVEELLSSASSLNFNSFTITVITLIN"

gene 7204..12342  
/gene="PFB0315w"  
7204..12342  
CDS  
/note="predicted by glimmer"  
/codon\_start=-1  
/product="hypothetical protein"  
/protein\_id="AAC71852.1"  
/db\_xref="PID:3845151"  
/db\_xref="GI:3845151"  
/translation="MDSKRYKRYVYNGFTKPPRYRLNDKGIHKLRLRVE  
PRLYNVDKIPKKEIYINISANNIOVRVONNNDEKKEANATCVANKRYTLAK  
VHVNKNSNINIKIYPTIKCSNFKIKNNISLKSYTEKFKLSNFSIKNGCS  
HNDVYNEMDQKSEQLNDNIRKLDVYCFREDTITKTNISYNNKNSPKNDE  
IYNNDNNKNSSSSYCSYKINONNVNHTLKTFLKNSHTQNEOSIPLDGLQ  
NNNSATKFNHNIYNNNSLVYKSDGIDLNKKMKIETDNGIITLEKKHDEKY  
NIFLPLNDNSNNVITTCODKESRSTSEMIKFEKMKNEKKNILKMNFDVI  
KKITMAKELINSSTINMKVSYFNKSDDELPEKSESYGVYKRENOEDIVYIK  
NKRNNININDNINIKNDVSKNINHNKKKDDPPNNSAGLIDPDCRYV  
LEILNVOSSKRRKILTNHNSSDNCHSSDQNCSSDQNCSSDQNCSSDQNC  
QCDNSNCKKDEKRRKKIKKRNKKNKNNKNNKNNKNNKNNKNNKNNKNN  
OCDNADGDIRINNEDKONVIONENGFNKNNKNNKNNKNNKNNKNNKNN  
KKENKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN  
VRESCIKRKYFYFOFDSYMRKKTITDMOVEDLYNENKIKRLKGGHDVY  
TMLEDNSYASGLNGENINFSNNENCLNSIDCEKYPINNEODEELCSNLY  
KDIERRKNGKLFEEICVFRINERNGHEHLRNHNDYHINSEYONLKNKO  
STNPKCKDEMKSQTNLFYDNYNSVDITLELNKNCYOHINVEYONLKNKESYA  
AEMNGLVFRKIPILINSCYLLIKNEKVIICISTYNDIVKAIYKSKNERS  
LEKIYVFRKKREKREKNTLFRANLMEIFEKIGRVDCIIPNEDKOLOKAK  
KKGKKELEKEMKRRKTOFYDIEVEYIEKRYRVHLYAGRLQIYERKPK  
YTKTSFHLRLIYEKTEMLKKNYTHCIDISDKLEFLIKMKOHFNHILNSY  
ESSFNYOIKTMDMLYNLLKESAYQHNLGNVILILYKVLNKKRMALYRFSV  
YNNIYVSKKNAFAYTLFRVNSILVAYERIKSFISKLKFNVDVSYFCFMYRILY  
RILFGYLIRDNIRINIKVIEKNYRVLKILISDNHKNYAFLOKRYEONKE  
NMICDNLIVANNECNLDKIAIEKGINOICLIFRKRKEGLMAYFTLNGPOINTE  
REYICIRYCSIFVNLKIOKKYOHIFPOYFKTIOKNNKRLTHAKILOVYOK  
EKSVIDVQLDYKPYFQIKDLRIEYVYICVQNFYTLNRLKLLNLELKLHLYK  
OEQFKYTYNGISITKFEVHVDKRLMNTIREFRYIANDKRLKMKMKMKOMK  
EKIDVNDMDQVMEQMDVMEQMDVMEQMDVMEQMDVMEQMDVMEQMDVMEQ  
YKRYKDSNSASLIRKYPFLINSEISPDCTTAGKRYNQNK"

BASE COUNT 5446 a 1189 c 1354 g 4951 t  
ORIGIN

Query Match 8.3%; Score 20; DB 37; Length 12940;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 TTTTGTGATTAATAATT 216  
Db 12335 TTTTGTGATTAATAATT 12316

RESULT 13  
OC063091 16609 bp DNA MAM 10-FEB-1997  
LOCUS Oryctolagus cuniculus beta-globin-like gene cluster locus control  
DEFINITION region.  
ACCESSION U63091 L05833 L05835  
NID 91835272



```
REFERENCE 1 (bases 1 to 135686)
AUTHORS Wilkinson,J.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-1997) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
COMMENT On May 14, 1997 this sequence version replaced g1:1172969.
IMPORTANT: This sequence is the entire insert of clone 473J6. This
sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre chromosome X
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/chrX/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 473J6 is at 1 in this sequence. The true
right end of clone 473J6 is at 135686.
473J6 is from the library RPC13 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.

FEATURES
Source 1..135686
location/qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="473J6"
/clone_lib="RPC1-3"
1..93
/note="MER25 repeat: matches 93. .1 of consensus"
558..764
/note="MER25 repeat: matches 1504. .1297 of consensus"
1336..1555
/note="MER25 repeat: matches 1296. .1079 of consensus"
1384..1877
/note="AluX repeat: matches 297. .1 of consensus"
1889..2345
/note="MER25 repeat: matches 1057. .602 of consensus"
2533..3027
/note="MER25 repeat: matches 548. .49 of consensus"
3476..3594
/note="FLAM_A repeat: matches 128. .11 of consensus"
3736..4786
/note="THEIB-INTERNAL repeat: matches 107. .1167 of
consensus"
4790..5109
/note="THEIC repeat: matches 54. .368 of consensus"
5113..6250
/note="LI repeat: matches 3226. .2078 of consensus"
6241..6480
/note="LIMB7 repeat: matches 242. .2 of consensus"
6333..7040
/note="LI repeat: matches 5390. .4668 of consensus"
7044..7347
/note="AluSq repeat: matches 1. .303 of consensus"
7370..9247
/note="LI repeat: matches 4680. .2773 of consensus"
9236..9607
/note="AluDo repeat: matches 1. .302 of consensus"
9631..9663
/note="AluSc repeat: matches 36. .4 of consensus;
incomplete repeat"
9667..10620
/note="LI repeat: matches 3978. .3022 of consensus"
10657..10951
/note="AluDo repeat: matches 301. .12 of consensus"
11983..12263
/note="AluDo repeat: matches 3. .277 of consensus;
incomplete repeat"
13944..13983
repeat_region
/note="20 copies of 2 mer 90 & conserved"
13984..14273
/note="AluDo repeat: matches 293. .6 of consensus"
16530..16701
/note="MLTIF repeat: matches 348. .539 of consensus"
17006..17206
/note="MLTIF repeat: matches 541. .343 of consensus"
17212..17582
/note="THEIC repeat: matches 371. .1 of consensus"
17588..19176
/note="THEIB-INTERNAL repeat: matches 1574. .1 of
consensus"
19179..19547
/note="THEIC repeat: matches 369. .2 of consensus"
19552..19776
/note="MLTIF repeat: matches 339. .113 of consensus"
23101..23301
/note="MLT1-INTERNAL repeat: matches 494. .287 of
consensus"
23368..23477
/note="MER34 repeat: matches 506. .401 of consensus"
24098..24307
/note="MLTIF repeat: matches 480. .268 of consensus"
24303..24727
/note="MER39 repeat: matches 6. .409 of consensus"
24813..24974
/note="MLTIF repeat: matches 272. .98 of consensus"
24893..24977
/note="MLTIG repeat: matches 186. .101 of consensus"
25312..26884
/note="LI repeat: matches 5183. .3613 of consensus"
26886..27112
/note="LI repeat: matches 5166. .5390 of consensus"
26962..27853
/note="LIP2A repeat: matches 1. .891 of consensus"
29189..29318
/note="FLAM_C repeat: matches 1. .130 of consensus"
31274..31580
/note="AluSq repeat: matches 303. .1 of consensus"
32974..33269
/note="AluSc repeat: matches 1. .295 of consensus"
33593..33915
/note="MER44A repeat: matches 2. .315 of consensus"
33690..33911
/note="MER44B repeat: matches 283. .514 of consensus"
34675..34762
/note="MER34 repeat: matches 151. .56 of consensus"
35890..35972
/note="MER4B repeat: matches 539. .457 of consensus"
35973..36271
/note="LIR8 repeat: matches 691. .388 of consensus"
36364..36430
/note="LIR8 repeat: matches 317. .251 of consensus"
36562..36719
/note="LIR8 repeat: matches 165. .1 of consensus"
36983..37138
/note="LIR8 repeat: matches 146. .1 of consensus"
37139..37276
/note="MER4B repeat: matches 460. .323 of consensus"
37677..37887
/note="MER4B repeat: matches 213. .1 of consensus"
37978..38325
/note="match: STS L41748"
42104..42309
/note="LIPAS repeat: matches 685. .890 of consensus"
49491..49516
/note="13 copies of 2 mer 92 & conserved"
49686..50316
/note="LI repeat: matches 4487. .5147 of consensus"
50306..52141
/note="LI repeat: matches 3536. .5390 of consensus"
52002..52648
/note="LIMB6 repeat: matches 3. .675 of consensus"
```

```

repeat_region 52644. .53044
/note="MLR2B repeat: matches 1. .404 of consensus"
repeat_region 53047. .53090
/note="22 copies of 2 mer 82 & conserved"
repeat_region 53141. .53215
/note="MLR2G repeat: matches 317. .388 of consensus"
repeat_region 53159. .53215
/note="MLR2B repeat: matches 390. .444 of consensus"
repeat_region 53217. .53414
/note="MLR2 Internal repeat: matches 2. .199 of consensus"
repeat_region 53415. .53690
/note="AluX repeat: matches 19. .292 of consensus;
incomplete repeat"
repeat_region 53964. .54006
/note="MADE1 repeat: matches 38. .80 of consensus"
repeat_region 55614. .56278
/note="LIP27 repeat: matches 890. .212 of consensus"
repeat_region 56278. .56700
/note="L1 repeat: matches 3941. .4388 of consensus"
repeat_region 57395. .57763
/note="AluO repeat: matches 302. .138 of consensus;
incomplete repeat"
repeat_region 58184. .58516
/note="MLR2 Internal repeat: matches 2057. .2400 of
consensus"
repeat_region 58676. .59124
/note="MLR2 Internal repeat: matches 3467. .3925 of
consensus"
repeat_region 59158. .59555
/note="MLR2B repeat: matches 1. .390 of consensus"
repeat_region 59556. .59638
/note="MLR1-INTERNAL repeat: matches 1. .84 of consensus"
repeat_region 60888. .60871
/note="AluO repeat: matches 7. .178 of consensus;
incomplete repeat"
repeat_region 60898. .61206
/note="AluX repeat: matches 302. .1 of consensus"
repeat_region 61242. .61373
/note="AluO repeat: matches 169. .301 of consensus;
incomplete repeat"
repeat_region 63678. .63849
/note="LIME3 repeat: matches 44. .225 of consensus"
repeat_region 64020. .64331
/note="L1 repeat: matches 1737. .1426 of consensus"

Query Match 7.98; Score 19; DB 9; Length 135686;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 168 TCAGATAAATAAATAAT 186
Db 31748 TCAGATAAATAAATAAT 31766
|||||

```

```

RESULT 15
LOCUS HS849L7/c 128169 bp DNA PRI 16-DEC-1997
DEFINITION Human DNA sequence from PAC 849L7 on chromosome Xq21.
ACCESSION AL008987
NID 92695813
VERSION AL008987.1 GI:2695813
KEYWORDS Xq21.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 128169)
REFERENCE
AUTHORS Direct Submission
JOURNAL Submitted (10-DEC-1997) Chromosome X Project Group
(http://www.sanger.ac.uk/HGP/ChrX/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

```

## COMMENT

On Dec 18, 1997 this sequence version replaced g1:2624153. IMPORTANT: This sequence is the entire insert of clone 849L7. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above. This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre chromosome X mapping group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX/>. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The true left end of clone 849L7 is at 1 in this sequence. The true right end of clone 849L7 is at 128169. 849L7 is from the library RPC15 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>.

## FEATURES

```

source
1..128169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq21"
/clone="849L7"
/clone_lib="RPC15"
1..633
/note="L1 repeat: matches 2369. .1729 of consensus"
643..1012
/note="THE1C repeat: matches 371. .3 of consensus"
1015..1493
/note="THE1B-INTERNAL repeat: matches 1580. .1104 of
consensus"
1524..2138
/note="MER25 repeat: matches 2136. .1509 of consensus"
3996..4570
/note="L1 repeat: matches 4796. .5384 of consensus"
4428..5327
/note="LIMB5 repeat: matches 5. .923 of consensus"
6655..6964
/note="AluX repeat: matches 1. .301 of consensus"
9408..9557
/note="MIR repeat: matches 53. .203 of consensus"
9643..9803
/note="MER3 repeat: matches 208. .36 of consensus"
10354..10844
/note="L1 repeat: matches 3682. .4182 of consensus"
10942..12182
/note="L1 repeat: matches 4143. .5390 of consensus"
12032..12879
/note="LIP27 repeat: matches 1. .899 of consensus"
12925..13055
/note="L1 repeat: matches 2259. .2127 of consensus"
13186..14053
/note="L1 repeat: matches 1886. .1017 of consensus"
13439..14063
/note="MER25 repeat: matches 2135. .1510 of consensus"
14173..14239
/note="MER25 repeat: matches 1225. .1160 of consensus"
14981..15892
/note="MER25 repeat: matches 1509. .582 of consensus"
16086..16587
/note="MER25 repeat: matches 518. .10 of consensus"
16641..16807
/note="LIM27 repeat: matches 875. .712 of consensus"
16641..16877
/note="LIM1 repeat: matches 875. .642 of consensus"
17047..17333

```



```
repeat_region /note="Alusq repeat: matches 4. .298 of consensus"
17441. .17524
/note="L1MB4 repeat: matches 928. .841 of consensus"
repeat_region 17530. .17824
/note="Alusx repeat: matches 3. .301 of consensus"
repeat_region 17827. .18364
/note="L1MB4 repeat: matches 851. .312 of consensus"
repeat_region 18357. .19293
/note="L1 repeat: matches 4140. .3190 of consensus"
repeat_region 19292. .20023
/note="L1 repeat: matches 2855. .2115 of consensus"
repeat_region 20238. .20482
/note="L1 repeat: matches 2001. .1761 of consensus"
repeat_region 20621. .21019
/note="MER25 repeat: matches 2124. .1717 of consensus"
repeat_region 21083. .21889
/note="L1PA2 repeat: matches 891. .85 of consensus"
repeat_region 21889. .21995
/note="L1 repeat: matches 5224. .5329 of consensus"
repeat_region 24964. .25221
/note="MER43 repeat: matches 272. .1 of consensus"
repeat_region 25671. .26529
/note="L1PA7 repeat: matches 890. .1 of consensus"
repeat_region 26385. .30002
/note="L1 repeat: matches 5390. .1770 of consensus"
repeat_region 30042. .30534
/note="L1PA7 repeat: matches 391. .892 of consensus"
repeat_region 32040. .32338
/note="Alusq repeat: matches 3. .301 of consensus"
repeat_region 33035. .34687
/note="L1 repeat: matches 3719. .5579 of consensus"
repeat_region 34721. .35023
/note="Alusq repeat: matches 1. .303 of consensus"
repeat_region 35024. .35051
/note="L1PA15 repeat: matches 160. .422 of consensus"
repeat_region 35053. .35316
/note="L1PA15 repeat: matches 160. .422 of consensus"
repeat_region 35320. .35717
/note="MSTR repeat: matches 3. .426 of consensus"
repeat_region 35723. .36207
/note="L1PA13 repeat: matches 424. .903 of consensus"
repeat_region 38047. .38222
/note="MIR repeat: matches 244. .68 of consensus"
repeat_region 41527. .42223
/note="L1PA16 repeat: matches 904. .195 of consensus"
repeat_region 42224. .42446
/note="Alusx repeat: matches 224. .2 of consensus;
incomplete repeat"
repeat_region 42585. .42624
/note="MIR2E repeat: matches 395. .356 of consensus"
repeat_region 44330. .44412
/note="PTRS repeat: matches 764. .676 of consensus"
repeat_region 47514. .47549
/note="18 copies of 2 mer 81 & conserved"
repeat_region 48744. .49205
/note="MIR1C repeat: matches 462. .6 of consensus"
repeat_region 49656. .50084
/note="L1MB2 repeat: matches 896. .458 of consensus"
repeat_region 50090. .50416
/note="MER31 repeat: matches 326. .1 of consensus"
repeat_region 50458. .50809
/note="MIR1A1 repeat: matches 1. .365 of consensus"
repeat_region 52127. .52275
/note="L1 repeat: matches 988. .1140 of consensus"
repeat_region 52154. .52520
/note="MER25 repeat: matches 1515. .1886 of consensus"
repeat_region 52521. .52593
/note="MAD1 repeat: matches 1. .76 of consensus"
repeat_region 52587. .52854
/note="MER25 repeat: matches 1850. .2126 of consensus"
repeat_region 52664. .54457
/note="L1 repeat: matches 1436. .3231 of consensus"
repeat_region 54458. .54911

repeat_region /note="L1PA8 repeat: matches 455. .909 of consensus"
54908. .57071
/note="L1 repeat: matches 3202. .5390 of consensus"
repeat_region 56924. .57841
/note="L1PA15 repeat: matches 1. .904 of consensus"
repeat_region 61576. .61903
/note="MER31 repeat: matches 330. .1 of consensus"
repeat_region 62025. .62479
/note="L1MB4 repeat: matches 466. .1 of consensus"
repeat_region 62331. .63804
/note="L1 repeat: matches 5390. .3932 of consensus"
repeat_region 65549. .66400
/note="L1MB2 repeat: matches 877. .2 of consensus"
repeat_region 67663. .67883
/note="L1MC1 repeat: matches 1074. .858 of consensus"
repeat_region 67884. .68295
/note="MIR1B repeat: matches 390. .1 of consensus"
repeat_region 68306. .68488
/note="L1MC3 repeat: matches 873. .687 of consensus"
repeat_region 68502. .68693
/note="L1PA4 repeat: matches 190. .382 of consensus"
repeat_region 68693. .68898
/note="L1PA3 repeat: matches 210. .1 of consensus"
repeat_region 68755. .69549
/note="L1 repeat: matches 5388. .4592 of consensus"
repeat_region 69579. .69703
/note="L1 repeat: matches 4464. .4590 of consensus"
repeat_region 69702. .69777
/note="Alusx/g repeat: matches 121. .194 of consensus;
incomplete repeat"
repeat_region 69826. .70524
/note="L1 repeat: matches 4619. .5316 of consensus"
repeat_region 71906. .72797
/note="L1PA2 repeat: matches 891. .1 of consensus"
repeat_region 72648. .73072
/note="L1 repeat: matches 5390. .4968 of consensus"
repeat_region 73031. .73153

Query Match 7.9%; Score 19; DB 9; Length 128169;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 TTGATTATTAATTTCT 219
|||||
DB 89211 TTGATTATTAATTTCT 89193
```

Search completed: August 6, 1999, 11:43:13  
Job time: 6679 sec

*This Page Blank (uspto)*

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:23:03 ; Search time 3059.95 Seconds  
(without alignments)  
156.000 Million cell updates/sec

Title: us-09-049-696-16

Perfect score: 242  
Sequence: 1 GTTATTCTCCACAGACTC.....ATGTTATTAGACTTCTGT 242

Scoring table:

Searched: 2546578 seqs, 986266752 residues

Database :

EST:  
1: em\_est1:  
2: em\_est2:  
3: em\_est3:  
4: em\_est4:  
5: em\_est5:  
6: em\_est6:  
7: em\_est7:  
8: em\_est8:  
9: em\_est9:  
10: em\_est10:  
11: em\_est11:  
12: em\_est12:  
13: em\_est13:  
14: em\_est14:  
15: em\_est15:  
16: em\_est16:  
17: em\_est17:  
18: em\_est18:  
19: em\_est19:  
20: em\_est20:  
21: em\_est21:  
22: em\_est22:  
23: em\_est23:  
24: em\_est24:  
25: em\_est25:  
26: em\_est26:  
27: em\_est27:  
28: em\_est28:  
29: em\_est29:  
30: em\_est30:  
31: em\_est31:  
32: em\_est32:  
33: em\_est33:  
34: em\_est34:  
35: em\_est35:  
36: em\_est36:  
37: em\_est37:  
38: em\_est38:  
39: em\_est39:  
40: em\_est40:  
41: em\_est41:  
42: em\_est42:  
43: em\_est43:  
44: em\_est44:  
45: em\_est45:  
46: em\_est46:  
47: em\_est47:  
48: em\_est48:  
49: em\_est49:  
50: em\_est50:  
51: em\_est51:  
52: em\_est52:  
53: em\_est53:

54: em\_est22:  
55: em\_est23:  
56: em\_est24:  
57: em\_est25:  
58: em\_est26:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	242	100.0	490	34	AA508854	AA508854 n122h10.s
2	242	100.0	716	49	AI660234	AI660234 w668902.x
3	213	88.0	392	50	AI687981	AI687981 w477f06.x
4	199	82.2	734	49	AI660957	AI660957 w420d08.x
5	164	67.8	235	20	D25727	D25727 HUMG504094
6	158	64.5	501	51	AI721275	AI721275 as82h08.x
7	156	64.3	335	51	AI721121	AI721121 as73d08.x
8	149	61.6	378	31	AA296955	AA296955 EST112726
9	114	47.1	161	48	AI582072	AI582072 as96a11.x
10	27	11.2	297	49	AI641931	AI641931 vq50e03.y
11	23	9.5	431	30	AA238284	AA238284 my34909.x
12	23	9.5	418	35	AA596289	AA596289 vo26a08.x
13	23	9.5	552	37	AA688953	AA688953 vs04h07.x
14	23	9.5	646	37	AA691335	AA691335 vs14h06.x
15	23	9.5	421	37	AA691586	AA691586 vs12g06.x
16	23	9.5	386	37	AA692521	AA692521 vs59g08.x
17	23	9.5	482	37	AA711228	AA711228 vt70c11.x
18	23	9.5	488	37	AA734415	AA734415 vt75b06.x
19	23	9.5	388	37	AA734161	AA734161 vs19g06.x
20	23	9.5	463	39	AA839323	AA839323 vo40f07.x
21	19	7.9	465	30	AA240794	AA240794 nm17c03.x
22	19	7.9	581	36	C81050	C81050 C81050 Mous
23	19	7.9	441	39	AA827192	AA827192 ob53h12.s
24	18	7.4	407	24	N31273	N31273 yx66c02.r1
25	18	7.4	232	25	N98035	N98035 2093C3 czap
26	18	7.4	157	33	C23819	C23819 C23819 Dict
27	18	7.4	330	34	AA493060	AA493060 v167g10.r
28	18	7.4	585	34	AA520990	AA520990 aa70c01.s
29	18	7.4	474	34	AA521102	AA521102 aa70g01.s
30	18	7.4	380	36	AA629484	AA629484 M0000037.
31	18	7.4	322	38	AA737071	AA737071 nx96e11.s
32	18	7.4	171	39	AA827932	AA827932 od02a12.s
33	18	7.4	544	41	AI041534	AI041534 cv82b07.x
34	18	7.4	528	41	AI042703	AI042703 uc76f07.x
35	18	7.4	509	41	AI042705	AI042705 uc76f09.x
36	18	7.4	528	41	AI042705	AI042705 uc76f09.x
37	18	7.4	652	44	AI316951	AI316951 u124f10.y
38	18	7.4	130	45	AI383804	AI383804 tc98f02.x
39	18	7.4	545	46	AI413216	AI413216 mc28g05.x
40	18	7.4	388	48	AI613951	AI613951 v167g10.y
41	18	7.4	421	50	AI692964	AI692964 w689h04.x
42	17	7.0	239	20	T37962	T37962 EST103279.S
43	17	7.0	380	20	T50928	T50928 yb88e01.r1
44	17	7.0	457	21	T86130	T86130 yd84c05.s1
45	17	7.0	460	22	R35323	R35323 y965b02.r1

#### ALIGNMENTS

RESULT 1  
LOCUS AA508854 490 bp mRNA  
DEFINITION n122h10.s1 NCI\_CGAP\_Co4 Homo sapiens cDNA clone IMAGE:968803 3',  
ACCESSION AA508854  
NID 92246357  
VERSION AA508854.1 GI:2246357

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 490)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397407.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert length: 856 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 404.

## FEATURES

## SOURCE

1. 490  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:968803"  
/clone\_lib="NCI-CGAP-Co4"  
/sex="pooled"  
/tissue\_type="colon"  
/lab\_host="DH10B"  
/note="Vector: pRT3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from pooled colon  
tumor tissue, and was then primed with a Not I - oligo(dT)  
primer. Double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pRT3 vector.  
This library is not normalized. Library constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 155 a 82 c 89 g 164 t

Query Match 100.0%; Score 242; DB 34; Length 490;  
Best Local Similarity 100.0%; Pred. No. 4.6e-104;

Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTATTCTCTCCACAGACTCCGCCAGACACCTAGCTCTGATGAAGCGTCTGCTTGG 60  
DB 408 GTTATTCTCTCCACAGACTCCGCCAGACACCTAGCTCTGATGAAGCGTCTGCTTGG 349  
QY 61 TCCATAATTCATATCAACAGCACCATTCCGCGATTCACATTTTAAATATATGTGGA 120  
DB 348 TCCATAATTCATATCAACAGCACCATTCCGCGATTCACATTTTAAATATATGTGGA 289  
QY 121 GTGATAGGAGAACTGAGCTGTCATATAGCTTAGGGCTGAATTTTGTCAATATAA 180  
DB 288 GTGATAGGAGAACTGAGCTGTCATATAGCTTAGGGCTGAATTTTGTCAATATAA 229  
QY 181 ATAAATATTCATCTCTTTTATATATAAATTTCTAAATGATTTTATAGACTTCT 240  
DB 228 ATAAATATTCATCTCTTTTATATATAAATTTCTAAATGATTTTATAGACTTCT 169  
QY 241 GT 242  
DB 168 GT 167

RESULT 2  
LOCUS A1660234/c  
DEFINITION w6860234.1 Soares.Dieckgraefe\_colon\_NHCD Homo sapiens CDNA clone  
IMAGE:2346290 3' similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA  
sequence.  
ACCESSION A1660234  
NID 94763804  
VERSION A1660234.1 GI:4763804  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 716)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949219.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 387.

## FEATURES

## SOURCE

1. 716  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2346290"  
/clone\_lib="Soares.Dieckgraefe.colon\_NHCD"  
/tissue\_type="colonic mucosa from 3 patients with Crohn's  
disease"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pRT3D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTCCATTCGAGAGTGGAGCGCGCGCTTTTATTTTATTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pRT3 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieck@im.wustl.edu); colonic mucosa represents a range of  
disease involvement from moderate to severe Crohn's  
disease; samples include both perforating (fistulas) and  
non-perforating samples. Library constructed by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 220 a 123 c 140 g 233 t

Query Match 100.0%; Score 242; DB 49; Length 716;  
Best Local Similarity 100.0%; Pred. No. 3.8e-104;

Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTATTCTCTCCACAGACTCCGCCAGACACCTAGCTCTGATGAAGCGTCTGCTTGG 60  
DB 405 GTTATTCTCTCCACAGACTCCGCCAGACACCTAGCTCTGATGAAGCGTCTGCTTGG 346  
QY 61 TCCATAATTCATATCAACAGCACCATTCCGCGATTCACATTTTAAATATATGTGGA 120  
DB 345 TCCATAATTCATATCAACAGCACCATTCCGCGATTCACATTTTAAATATATGTGGA 286  
QY 121 GTGATAGGAGAACTGAGCTGTCATATAGCTTAGGGCTGAATTTTGTCAATATAA 180  
DB 285 GTGATAGGAGAACTGAGCTGTCATATAGCTTAGGGCTGAATTTTGTCAATATAA 226  
QY 181 ATAAATATTCATCTCTTTTATATATAAATTTCTAAATGATTTTATAGACTTCT 240

DB 225 ATRAAATCATCATCCTTTTGTGATATATAAATTTCTAAATGATTTAGACTTCT 166  
OY 241 GT 242  
DB 165 GT 164

RESULT 3  
A1687981 392 bp mRNA EST 27-MAY-1999  
LOCUS waf7f06.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
DEFINITION IMAGE:2302211 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA  
sequence.  
ACCESSION A1687981  
NID A1687981  
VERSION 94899275  
KEYWORDS A1687981.1 GI:4899275  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 392)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2947973.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gldco.  
Location/Qualifiers  
1..392  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2302211"  
/clone\_11b="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site: 1. Not I; Site: 2. Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NDHL19W, testis NHT, and B-cell  
NCI-CGAP GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 125 a 69 c 67 g 131 t  
ORIGIN

Query Match 88.0%; Score 213; DB 50; Length 392;  
Best Local Similarity 100.0%; Pred. No. 2.1e-90;  
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 CTCGCCGAGACACCTAGTCTGATGAAGCTGCTCTGCTCTTAATATTCATATCA 77  
DB 355 CTCGCCGAGACACCTAGTCTGATGAAGCTGCTCTGCTCTTAATATTCATATCA 296  
OY 78 ACAGACACATTCCTGGCANTCATTTTAAATATGGAAGGAGGAGGAACCTGC 137  
DB 235 ACAGACACATTCCTGGCANTCATTTTAAATATGGAAGGAGGAGGAACCTGC 236  
OY 138 AGCTGTCAATAGCTAGGGCTGAATTTTGTGAGATTAATTAATCATTCATCTT 197  
DB 235 AGCTGTCAATAGCTAGGGCTGAATTTTGTGAGATTAATTAATCATTCATCTT 176

OY 198 TTTTGTGATATAAATTTCTAAATGATTT 230  
DB 175 TTTTGTGATATAAATTTCTAAATGATTT 143

RESULT 4  
A1660957 734 bp mRNA EST 10-MAY-1999  
LOCUS w120d08.x1 Soares\_Dieckgraefe\_colon\_NHUC Homo sapiens cDNA clone  
DEFINITION IMAGE:2351151 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA  
sequence.  
ACCESSION A1660957  
NID 94764540  
VERSION A1660957.1 GI:4764540  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 734)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980033.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gldco.  
High quality sequence stop: 477.  
Location/Qualifiers  
1..734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="X"  
/clone="IMAGE:2351151"  
/clone\_11b="Soares\_Dieckgraefe\_colon\_NHUC"  
/tissue\_type="colonic mucosa from 5 ulcerative colitis  
patients"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site: 1. Not I; Site: 2. Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAAGGAGCGCGCGCTTTTATTTTATTTT 3']  
TGTACCAATCTGAAGGAGCGCGCGCTTTTATTTTATTTT 3']  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieck@km.wustl.edu); colonic mucosa represents a range of  
disease involvement from mild cryptitis to severe  
ulceration, fibrosis, and degeneration. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 200 a 143 c 159 g 228 t  
ORIGIN

Query Match 82.2%; Score 199; DB 49; Length 734;  
Best Local Similarity 100.0%; Pred. No. 5.6e-84;  
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTTATTCCTCCACAGACTCCGCCAGAGACACCTAGTCTGTAAGAGTGTCTCTTG 60  
DB 201 GTTATTCCTCCACAGACTCCGCCAGAGACACCTAGTCTGTAAGAGTGTCTCTTG 142  
OY 61 TCTTAATATTCATATCAACAGACCATTCCTGGCATTTCACATTTTAAATATATGTGGA 120  
DB 141 TCTTAATATTCATATCAACAGACCATTCCTGGCATTTCACATTTTAAATATATGTGGA 82  
OY 121 GTGATAGGAGACTGCAGCTGTCAATAGCCTAGGCTGAATTTTGTGACATTAATAA 180

```

Db      81 GTGGATAGAGAGACTGCAGCTGTCATAGCTAGGCGCTGAATTTTGTGCAGTAATAATA 22
Qy      181 ATAAATCATTCATCCCTTT 199
Db      21 ATAAATCATTCATCCCTTT 3

RESULT  5
D25727 255 bp mRNA EST 30-NOV-1995
LOCUS   HUMGS04094 Human colon mucosa Homo sapiens cDNA clone cm2037 3',
DEFINITION mRNA sequence.
ACCESSION D25727
NID       9500422
VERSION   D25727.1 GI:500422
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Okubo, K., Yoshii, J., Yokouchi, H., Kameyama, M. and Matsubara, K.
AUTHORS   1 (bases 1 to 255)
TITLE     Global analysis of gene expression in colon mucosa: a large scale
COMMENT   random cDNA sequencing analysis
           Unpublished (1994)

FEATURES
Source
Location/Qualifiers
1..255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cm2037"
/note="Adult male, tissue-type = colon mucosa"
BASE COUNT 76 a 52 c 39 g 80 t 8 others
ORIGIN

Query Match 67.8%; Score 164; DB 20; Length 255;
Best Local Similarity 100.0%; Pred. No. 2.5e-67;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 CCTCCACAGACTCCGCCAGAGACAGACCTAGTCTGTGTAAGGTCCTGCTGCTGCTAAT 67
Db      49 CCTCCACAGACTCCGCCAGAGACAGACCTAGTCTGTGTAAGGTCCTGCTGCTGCTAAT 108
Qy      68 ATTATATCAACACACACCATTCCTGGCATTCACATTTTAAATAATTAATGTGAAGTGATA 127
Db      109 ATTATATCAACACACACCATTCCTGGCATTCACATTTTAAATAATTAATGTGAAGTGATA 168
Qy      128 GGAGAACTGACGTGTCATATAGCCTAGAGGCTGAATTTTGTGAC 171
Db      169 GGAGAACTGACGTGTCATATAGCCTAGAGGCTGAATTTTGTGAC 212

RESULT  6
A1721275 501 bp mRNA EST 10-JUN-1999
LOCUS   as82h08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
DEFINITION IMAGE:2335263 3' similar to SW:EGIC.BOVIN P54281 EPITHELIAL
           CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.
ACCESSION A1721275
NID       95038531
VERSION   A1721275.1 GI:5038531
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 501)
AUTHORS   Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
           Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marz, M.,
           Martin, J., Moore, B., Schellenberg, R., Steptoe, M., Tan, F.,
           Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE     WashU-NCI human EST Project
COMMENT   Unpublished (1997)
           On Jun 22, 1998 this sequence version replaced gi:3246918.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -400P from Gdbco
High quality sequence stop: 394.

FEATURES
Source
Location/Qualifiers
1..501
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="Xp11.2"
/clone="IMAGE:2335263"
/clone_lib="Barstead colon HPLRB7"
/sex="male"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/note="Organ: colon; Vector: pRT3D-Pac (Pharmacia) with a
modified polynuker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer [5'
TGTTCAGAACTGAGTGAAGTGGAGGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[5' ATTTCACCTAGTAT 3' and 5' ATTCTAGT 3'], digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pRT3 vector. Library constructed by Bob
Barstead."
BASE COUNT 158 a 83 c 106 g 154 t
ORIGIN

Query Match 65.3%; Score 158; DB 51; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.2e-64;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GTTATTCCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAAGACGCTCTCTCTG 60
Db      214 GTTATTCCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAAGACGCTCTCTCTG 155
Qy      61 TCCTAATATTCATATCAACAGACCATTCCTGGCATTCACATTTTAAATAATTAATGTGAA 120
Db      154 TCCTAATATTCATATCAACAGACCATTCCTGGCATTCACATTTTAAATAATTAATGTGAA 95
Qy      121 GTGATAGGAGAACTGACGTGTCATATAGCTTAGGGCT 158
Db      94 GTGATAGGAGAACTGACGTGTCATATAGCTTAGGGCT 57

RESULT  7
A1721121 335 bp mRNA EST 10-JUN-1999
LOCUS   as73d08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
DEFINITION IMAGE:2334351 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA
           sequence.
ACCESSION A1721121
NID       95038377
VERSION   A1721121.1 GI:5038377
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```



SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 161)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S., Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maitra, M., Mattin, J., Moore, B., Schellenberg, K., Sepcoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Mar 19, 1997 this sequence version replaced gi:1900677.

Contact: Wilson R.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -400P from Gibco  
High quality sequence stop: 115.  
Location/Qualifiers  
1. 161  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2173436"  
/clone\_1b="Barstead colon HPRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGATCTGAGTGGAGCGCGCCCTTTTGTGATTATATAA 3']; double-stranded cDNA was ligated to Eco RI adaptors [5' ATTCTAGTATAT 3' and 5' ATTCTAGT 3'], digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library constructed by Bob Barstead."

BASE COUNT 55 a 24 c 29 g 53 t  
ORIGIN

Query Match 47.1%; Score 114; DB 48; Length 161;  
Best Local Similarity 100.0%; Pred. No. 9.3e-44;  
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 CATTTTAAATTTATGAGATGATGAGAACTGCAGCTGCATATGCTAGGCTG 159  
|||||  
Db 116 CATTTTAAATTTATGAGATGATGAGAACTGCAGCTGCATATGCTAGGCTG 57  
|||||

QY 160 AATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 213  
|||||  
Db 56 AATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3  
|||||

RESULT 10  
LOCUS A1641931 297 bp mRNA EST 29-APR-1999  
DEFINITION v950603.y1 Barstead bowel MFLRB9 Mus musculus cDNA clone  
IMAGE:1097692 5' similar to TR:088826 O88826 GOR-5 PROTEIN.; mRNA  
sequence.  
ACCESSION A1641931  
NID 94720406  
VERSION A1641931.1 GI:4720406  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 297)

AUTHORS Maitra, M., Hillier, L., Kucaba, T., Mattin, J., Beck, C., Wylie, T., Underwood, K., Sepcoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
TITLE WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948705.

Contact: Maitra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
NCI:603924  
This read is a RESEQUENCE of a previously sequenced mouse clone  
this read has been verified (found to hit its original self in the correct orientation)  
putative full length read  
vector to vector length is 481  
possible reversed clone: similarity on wrong strand  
Seq primer: -40RP from Gibco  
High quality sequence stop: 288.  
Location/Qualifiers  
1. 297  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1097692"  
/clone\_1b="Barstead bowel MFLRB9"  
/tissue\_type="bowel"  
/dev\_stage="8 weeks"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGATCTGAGTGGAGCGCGCCCTTTTGTGATTATATAA 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATCGGATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Source irradiated bowel harvested 72 hours after irradiation (1400 Gys). Library constructed by Bob Barstead."

BASE COUNT 80 a 88 c 58 g 71 t  
ORIGIN

Query Match 11.2%; Score 27; DB 49; Length 297;  
Best Local Similarity 100.0%; Pred. No. 0.0046;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATCAACAGCACCATTCCTGCGATTCAC 100  
|||||  
Db 262 ATCAACAGCACCATTCCTGCGATTCAC 288  
|||||

RESULT 11  
LOCUS AA238284 431 bp mRNA EST 03-MAR-1997  
DEFINITION my34909.r1 Barstead mouse pooled organs MFLRB4 Mus musculus cDNA  
clone IMAGE:697792 5' similar to TR:01844066 G1184066  
CALCIUM-ACTIVATED CHLORIDE CHANNEL.; mRNA sequence.  
ACCESSION AA238284  
NID 91862488  
VERSION AA238284.1 GI:1862488  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.



TITLE	JOURNAL	COMMENT
Thelsting,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
The WashU-HHMI Mouse EST Project		
Unpublished (1996)		
On Sep 12, 1996 this sequence version replaced gi:1290647.		
CONTACT: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mousestewartson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. MGI:582574 Seq primer: -28ml3 rev2 Et from Amersham High quality sequence stop: 418.		
FEATURES		
SOURCE		
1. 418 Location/Qualifiers		
/organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:10510998" /clone_11b="Barstead mouse irradiated colon MPLRB7" /dev_stage="8 weeks" /db_host="DHI0B" /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site:1: EcoRI; Site:2: NotI. Tissue obtained from 8 week old mouse. Colon was harvested 72 hours after irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - 0150(gt) primer [5'TGTTCAGATCTGAGTGGAGCGAGCGCCGCCCTTTTCTTTTCTTTTCTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATCGCATCGCTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library constructed by Bob Barstead."		
BASE COUNT		
111 a 96 c 102 g 109 t		
ORIGIN		
Query Match 9 5%; Score 23; DB 35; Length 418; Best Local Similarity 100.0%; Prid. NO. 0.29; Matches 23; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;		
OY 74 ATCAACAGCACCATTCTCTGCAT 96 		
DB 210 ATCAACAGCACCATTCTCTGCAT 232		
RESULT 13 AA688953		
LOCUS		
DEFINITION		
AA688953 552 bp mRNA EST 12-DEC-1997 VS04h07.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA clone IMAGE:113727 5' similar to SW:ECIC_BOVIN P54281 EPITHELIAL CHLORIDE CHANNEL PROTEIN ; mRNA sequence.		
ACCESSION		
AA688953		
IID		
92678382		
VERSION		
AA688953.1 GI:2678382		
KEYWORDS		
EST.		
SOURCE		
house mouse. Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
ORGANISM		
REFERENCE		
AUTHORS		
Marra,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelsting,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE		
The WashU-HHMI Mouse EST Project		
JOURNAL		
Unpublished (1996)		
On Nov 6, 1997 this sequence version replaced gi:930439.		
COMMENT		

Contact: Marra M/Mouse EST Project  
 WashU-HM1 Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:618549  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 509.  
 Location/Qualifiers

## FEATURES

source

1..552  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /map="13"  
 /clone\_image="1137277"  
 /clone\_lib="Barstead mouse irradiated colon MPLRB7"  
 /dev\_stage="8 weeks"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained  
 from 8 week old mouse. Colon was harvested 72 hours after  
 irradiation with 1400 Gys. 1st strand cDNA was primed  
 with a Not I - oligo(dT) primer  
 [5'TGTACGATCTGACGTGGAGCGCCGCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors [AATTCGATCTTG], digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT73  
 vector. Library constructed by Bob Barstead."

## BASE COUNT

152 a 148 c 122 g 130 t

## ORIGIN

Query Match 9.5%; Score 23; DB 37; Length 552;  
 Best Local Similarity 100.0%; Pred. No. 0.25;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ATCAACAGCACCATCTCGGCAT 96  
 ||||||||||||||||||||  
 DB 397 ATCAACAGCACCATCTCGGCAT 419

RESULT 14  
 AA691335 646 bp mRNA EST 16-DEC-1997  
 LOCUS  
 DEFINITION vs14106.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus CDNA  
 clone IMAGE:1138235 5' similar to SW:EDIC\_BOVIN P54281 EPITHELIAL  
 CHLORIDE CHANNEL PROTEIN ; mRNA sequence.

ACCESSION  
 NID 92692271  
 VERSION AA691335.1 GI:2692271  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 (bases 1 to 646)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The WashU-HM1 Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1397564.

Contact: Marra M/Mouse EST Project  
 WashU-HM1 Mouse EST Project  
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:619507  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 491.  
 Location/Qualifiers

## FEATURES

source

1..646  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /map="17q21; 2"  
 /clone\_image="1138235"  
 /clone\_lib="Barstead mouse irradiated colon MPLRB7"  
 /dev\_stage="8 weeks"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained  
 from 8 week old mouse. Colon was harvested 72 hours after  
 irradiation with 1400 Gys. 1st strand cDNA was primed  
 with a Not I - oligo(dT) primer  
 [5'TGTACGATCTGACGTGGAGCGCCGCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors [AATTCGATCTTG], digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT73  
 vector. Library constructed by Bob Barstead."

## BASE COUNT

171 a 176 c 153 g 146 t

## ORIGIN

Query Match 9.5%; Score 23; DB 37; Length 646;  
 Best Local Similarity 100.0%; Pred. No. 0.24;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ATCAACAGCACCATCTCGGCAT 96  
 ||||||||||||||||||||  
 DB 584 ATCAACAGCACCATCTCGGCAT 606

RESULT 15  
 AA691586 421 bp mRNA EST 16-DEC-1997  
 LOCUS  
 DEFINITION vs12906.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus CDNA  
 clone IMAGE:1138042 5' similar to SW:EDIC\_BOVIN P54281 EPITHELIAL  
 CHLORIDE CHANNEL PROTEIN ; mRNA sequence.

ACCESSION  
 NID 92692524  
 VERSION AA691586.1 GI:2692524  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 (bases 1 to 421)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The WashU-HM1 Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1405150.

Contact: Marra M/Mouse EST Project  
 WashU-HM1 Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.lnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 391.

Location/Qualifiers

# FEATURES

source

1. 421

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:1138042"

/clone\_lib="Barstead mouse irradiated colon MPLRB7"

/dev\_stage="8 weeks"

/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained

from 8 week old mouse. Colon was harvested 72 hours after

irradiation with 1400 Gys. 1st strand cDNA was primed

with a Not I - oligo(dT) primer

15'GTTCAGATCTGAGTGGAGCGGCGCCCTTTTCTTTTCTTTTCTTTT

T 3'; double-stranded cDNA was ligated to Eco RI

adaptors [AATCGGATCCTG], digested with Not I and cloned

into the Not I and Eco RI sites of the modified pT73

vector. Library constructed by Bob Barstead.

BASE COUNT 119 a 98 c 89 g 115 t

ORIGIN

Query Match 9.5%; Score 23; DB 37; Length 421;

Best Local Similarity 100.0%; Pred. No. 0.29; Mismatches 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ATCAACAGCACCATTCTGGCAT 96

Db 173 ATCAACAGCACCATTCTGGCAT 195

Search completed: August 6, 1999, 09:23:06

Job time: 6257 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:14:56 ; Search time 3264.21 Seconds

(without alignments)  
235.780 Million cell updates/sec

Title: US-09-049-696-16

Perfect score: 242  
Sequence: 1 GTTATTCTCCTCAGACTC.....ATGTAATTTAGACTTCTGT 242

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: \*  
1: gb\_da1:\*  
2: gb\_da2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pi1:\*  
8: gb\_pi2:\*  
9: gb\_pi3:\*  
10: gb\_pi4:\*  
11: gb\_pi5:\*  
12: gb\_pi6:\*  
13: gb\_pi7:\*  
14: gb\_pi8:\*  
15: gb\_pi9:\*  
16: gb\_pi10:\*  
17: gb\_pi11:\*  
18: gb\_pi12:\*  
19: gb\_pi13:\*  
20: gb\_pi14:\*  
21: gb\_pi15:\*  
22: gb\_pi16:\*  
23: gb\_pi17:\*  
24: gb\_pi18:\*  
25: gb\_pi19:\*  
26: gb\_pi20:\*  
27: gb\_pi21:\*  
28: gb\_pi22:\*  
29: gb\_pi23:\*  
30: gb\_pi24:\*  
31: gb\_pi25:\*  
32: gb\_pi26:\*  
33: gb\_pi27:\*  
34: gb\_pi28:\*  
35: gb\_pi29:\*  
36: gb\_pi30:\*  
37: gb\_pi31:\*  
38: gb\_pi32:\*  
39: gb\_pi33:\*  
40: gb\_pi34:\*  
41: gb\_pi35:\*  
42: gb\_pi36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	230	95.0	3311	11	AF039400	AF039400 Homo sapi

2	228.4	94.2	35278	11	AF039401	AF039401 Homo sapi
3	213.4	88.2	2826	42	AF127036	AF127036 Homo sapi
4	166.8	68.9	878	5	I95746	I95746 Sequence 8
5	92	38.0	2937	12	AB017156	AB017156 Mus muscu
6	43.8	18.1	256172	35	AC005139	AC005139 Plasmodiu
7	43.8	18.1	284972	35	AC005505	AC005505 Plasmodiu
8	41	16.9	138339	34	PFMAL1P1	AL031744 Plasmodiu
9	41	16.9	16343	36	AMEGENOM	L06178 Apis mellif
10	41	16.9	2950	36	AMEFTCOX	M33409 Honey bee m
11	40.8	16.9	107702	7	AF117L22	AL035527 Arabidops
12	40.6	16.8	53778	8	T1J1	AF128393 Arabidops
13	39.2	16.6	80891	8	ATAC005693	AC005693 Arabidops
14	39.6	16.4	117752	11	AC004743	AC004743 Homo sapi
15	39.4	16.3	110000	34	CEY39E4_1	Continuation (2 of
16	39.4	16.3	110000	34	PFMAL4P1_0	AL034357 Plasmodiu
17	39.2	16.2	8212	7	ALRI1BPT	X75653 A. longia pla
18	39.2	16.2	261002	34	CEY41C4	AL022278 Caenorhab
19	39	16.1	224100	35	AC007528	AC007528 Homo sapi
20	39	16.1	275	36	TRBVSGERC2	M30741 T. brucei va
21	39	16.1	232	36	TRBVSGERT2	M30743 T. brucei va
22	38.6	16.0	29324	1	BS16829XB	AJ222587 Bacillus
23	38.6	16.0	3215	1	BSCHVEGE	Z29584 B. subtilis
24	38.6	16.0	3215	1	BSU05345	U05345 Bacillus su
25	38.6	16.0	208230	1	BSUB0008	Z99111 Bacillus su
26	38.4	15.9	175374	34	PFMAL1P4	AL031747 Plasmodiu
27	38.4	15.9	114226	35	AC004710	AC004710 Plasmodiu
28	38.4	15.9	104992	35	AC005504	AC005504 Plasmodiu
29	38	15.7	98957	8	ATAC004136	AC004136 Arabidops
30	38	15.7	219200	35	AC006279	AC006279 Plasmodiu
31	38	15.7	114736	36	PFMAL3P3	Z98547 Plasmodiu
32	38	15.7	16741	36	PESC03095	AL010164 Plasmodiu
33	37.8	15.6	36454	34	AC006670	AC006670 Caenorhab
34	37.8	15.6	274626	34	AC006903	AC006903 Caenorhab
35	37.8	15.6	227175	35	AC007253	AC007253 Homo sapi
36	37.8	15.6	678	36	AMEFTMICOA	M77208 Apis andren
37	37.6	15.5	4587	17	EVU67964	U67964 Ectromella
38	37.4	15.5	110680	8	AC006259	AC006259 Arabidops
39	37.4	15.5	160457	42	AC006016	AC006016 Homo sapi
40	37.4	15.5	47042	42	AC007031	AC007031 Homo sapi
41	37.2	15.4	119704	7	OSCHLEPLX	Z67753 O. sinensis
42	37.2	15.4	436	8	AF114934	AF114934 Saccharom
43	37.2	15.4	40997	17	CV41KBPL	X94355 Cowpox viru
44	37.2	15.4	52283	17	CVGR190	Y11842 Cowpox viru
45	37.2	15.4	4618	17	CVOR115L	X83621 Cowpox viru

## ALIGNMENTS

RESULT 1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS	AF039400	3311 bp	mrna		
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hclcal) mRNA, complete cds.				
ACCESSION	AF039400				
VERSION	94009457				
KEYWORDS	AF039400.1 GI:4009457				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Gruber/A.D., Elble/R.C., Ji/H.L., Schreuer/R.D., Fuller/C.M. and Pauli/B.U.				
TITLE	Genomic cloning, molecular characterization, and functional analysis of human hclcal, the first human member of the family of Ca2+-activated Cl- channel proteins				
JOURNAL	Genomics 54 (2), 200-214 (1998)				
MEDLINE	99047526				
REFERENCE	2 (bases 1 to 3311)				
AUTHORS	Gruber/A.D., Elble/R. and Pauli/B.U.				
TITLE	Direct Submission				

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

## FEATURES

## Source

Location/Qualifiers

1. .3311

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="1p22-p31"

/tissue\_type="small intestine"

1. .3311

/gene="hCLCA1"

352. .3096

/gene="hCLCA1"

/note="transmembrane glycoprotein"

/codon\_start=1

/product="calcium-dependent chloride channel-1"

/protein\_id="AAC95428.1"

/db\_xref="PID:94009458"

/db\_xref="GI:4009458"

/translation="MGPKRSVFLILHLLEGALSNLIQLNNNGEIVVAIDPNV  
EDERLIQIKDMVTQASLYLEPATKRYFKVALLPEYMTKADYVPRLEYKNA  
DYLVAESTPNDDEPTEDMGNGCKGERGRIHLTPDLAKGLAEYPOGKAFVEMAH  
LRMGVDEYNDEKYEYLSNGRIQAVCSAGITGVVVKCOGSGCYTRCFNVTGL  
YEKGEFVLOSROTEKASIMFAQHDVSIYECITEONHNKEAPNKNONCNLSRMEVI  
RDSDEKRTTPTPTOPNPPTFSILOIGRICIVLDKSGMATGRMLRLNAGOLPL  
LOTVELGSMVGMWTFDSDAHVOSSELIQINSQSDPTLAKRPAASGGTSICGLRRA  
FTVIRKKYPTDGSSEIVLITDGEDNTISCFNPKYKOSGCIHTVALGSRADLELSK  
MTGGLQTVASDOVONNGLIDAFGLSSGNGAVSOSIQLESGLTLQNSQMNNTVIY  
DSVIGKDTLFLITWTPOPIILMDPSQKOGGVNDKTMALQIPIGIRKWTWY  
SLQASQDTLITVTSRASNAATLPITVTSKTNKTSKPSPLVYANROGASPLRA  
SVTALIESVNGKTVLLELDNAGADATKDGVSRYFTYDINGRYSVKRALGVN  
AARRVTPQOSGALYIPGMIENDEIOWNPPEINPKPDYDGHGTAHKTIIRISITLD  
DVPNAPIDPLEPPOGTTDLNLAIEHGSLINLTWTAPOGDYQHVCFSTRSSGGSFVAS  
LBDKNESLQVNTTALIPKEANSEVFLFKENITFEENGDTLFAIQADVVDLKSEI  
SNIAVSLFIPPTPTPTSPDTSAPCPNHNINSTIPIGHIHLKIMMKWIGELQLSIA

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

## Query Match

Best Local Similarity 95.0%; Score 230; DB 11; Length 3311;

Matches 241; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
OY 1 GTTATTCCTCCAGACGACGCGCAGACGACCTAGTCTGTAAGTCTGCTCTG 60
DB 2943 GTTATTCCTCCAGACGACGCGCAGACGACCTAGTCTGTAAGTCTGCTCTG 3002
OY 61 TCCTAATATTCATATCAACAGACGACCATTCCTGCAATTTAAATATATG 120
DB 3003 TCCTAATATTCATATCAACAGACGACCATTCCTGCAATTTAAATATATG 3062
OY 121 GTGATAGAGAACTGCGAGCTGTCAATAGCCCTGGAATTTTGCAGATAAATA 180
DB 3063 GTGATAGAGAACTGCGAGCTGTCAATAGCCCTGGAATTTTGCAGATAAATA 3122
OY 181 AATAATCATTCCTTTTGTGATTAATAATTTCTAATAAGCTATTTAGACTTCT 240
DB 3123 AATAATCATTCATCC-TTTTTTGTATTAATAATTTCTAATAATTTTAGACTTCT 3181
OY 241 GT 242
DB 3182 GT 3183
```

RESULT 2  
AF039401  
LOCUS AF039401 35278 bp DNA PRI 15-DEC-1998  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.  
ACCESSION AF039401  
MID 94009459  
VERSION AF039401.1 GI:4009459

## KEYWORDS

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 35278)

Gruber,A.D., Edle,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and

Pauli,B.U.

Genomic cloning, molecular characterization, and functional

analysis of human CLCA1, the first human member of the family of

Ca2+-activated Cl- channel proteins

Genomics 54 (2), 200-214 (1998)

99047526

2 (bases 1 to 35278)

Gruber,A.D., Edle,R. and Pauli,B.U.

Direct Submission

Submitted (19-DEC-1997) Department of Pathology, College of

Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

Location/Qualifiers

1. .35278

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="1p22-p31"

1. .1617

/gene="hCLCA1"

1. .33522

/gene="hCLCA1"

767. .1265

/rpt\_family="L1 transposable element"

1587. .1596

/rpt\_type="dispersed"

/gene="hCLCA1"

/note="putative"

1618. .1708

/gene="hCLCA1"

/number=1

join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..24522,24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)

/product="calcium-dependent chloride channel-1"

1962. .2383

/gene="hCLCA1"

/number=2

join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..24522,24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)

/gene="hCLCA1"

/codon\_start=1

/product="calcium-dependent chloride channel-1"

/protein\_id="AAC95429.1"

/db\_xref="PID:94009460"

/db\_xref="GI:4009460"

/translation="MGPKRSVFLILHLLEGALSNLIQLNNNGEIVVAIDPNV  
EDERLIQIKDMVTQASLYLEPATKRYFKVALLPEYMTKADYVPRLEYKNA  
DYLVAESTPNDDEPTEDMGNGCKGERGRIHLTPDLAKGLAEYPOGKAFVEMAH  
LRMGVDEYNDEKYEYLSNGRIQAVCSAGITGVVVKCOGSGCYTRCFNVTGL  
YEKGEFVLOSROTEKASIMFAQHDVSIYECITEONHNKEAPNKNONCNLSRMEVI  
RDSDEKRTTPTPTOPNPPTFSILOIGRICIVLDKSGMATGRMLRLNAGOLPL  
LOTVELGSMVGMWTFDSDAHVOSSELIQINSQSDPTLAKRPAASGGTSICGLRRA  
FTVIRKKYPTDGSSEIVLITDGEDNTISCFNPKYKOSGCIHTVALGSRADLELSK  
MTGGLQTVASDOVONNGLIDAFGLSSGNGAVSOSIQLESGLTLQNSQMNNTVIY  
DSVIGKDTLFLITWTPOPIILMDPSQKOGGVNDKTMALQIPIGIRKWTWY  
SLQASQDTLITVTSRASNAATLPITVTSKTNKTSKPSPLVYANROGASPLRA  
SVTALIESVNGKTVLLELDNAGADATKDGVSRYFTYDINGRYSVKRALGVN  
AARRVTPQOSGALYIPGMIENDEIOWNPPEINPKPDYDGHGTAHKTIIRISITLD  
DVPNAPIDPLEPPOGTTDLNLAIEHGSLINLTWTAPOGDYQHVCFSTRSSGGSFVAS  
LBDKNESLQVNTTALIPKEANSEVFLFKENITFEENGDTLFAIQADVVDLKSEI  
SNIAVSLFIPPTPTPTSPDTSAPCPNHNINSTIPIGHIHLKIMMKWIGELQLSIA

exon 6667..6807  
/gene="hCLCA1"  
/number=3  
exon 6979..7126  
/gene="hCLCA1"  
/number=4  
exon 9696..9801  
/gene="hCLCA1"  
/number=5  
exon 15454..15631  
/gene="hCLCA1"  
/number=6  
exon 18591..18809  
/gene="hCLCA1"  
/number=7  
exon 19783..20010  
/gene="hCLCA1"  
/number=8  
exon 22253..22427  
/gene="hCLCA1"  
/number=9  
exon 24522..24628  
/gene="hCLCA1"  
/number=10  
exon 26644..26859  
/gene="hCLCA1"  
/number=11  
exon 27447..27708  
/gene="hCLCA1"  
/number=12  
exon 28766..28936  
/gene="hCLCA1"  
/number=13  
exon 31840..32079  
/gene="hCLCA1"  
/number=14  
exon 32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN

Query Match 94.4%; Score 228.4; DB 11; Length 35278;  
Best Local Similarity 99.2%; Pred. No. 2.8e-37;

Matches 240; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GTTTATCTCCACAGACCTCCGCGACAGACACTACTCTGATGAACGCTGCTCCTTG 60  
DB 33157 GTTTATCTCCACAGACCTCCGCGACAGACACTACTCTGATGAACGCTGCTCCTTG 33216  
QY 61 TCCTAATATTCATATCAACAGACCACTTCCTGCAATTTAAATTTATGTGAA 120  
DB 33217 TCCTAATATTCATATCAACAGACCACTTCCTGCAATTTAAATTTATGTGAA 33276  
QY 121 GTGATAGAGAACTGCACTGTCAATAGCCTAGGCGTGAATTTTGTGAGATTAATA 180  
DB 33277 GTGATAGAGAACTGCACTGTCAATAGCCTAGGCGTGAATTTTGTGAGATTAATA 33336  
QY 181 ATAAATCATTCATCTTTTGTGATTAATAATTTCTTAATTTGATTTAGACTTCT 240  
DB 33337 ATAAATCATTCATCTTTTGTGATTAATAATTTCTTAATTTGATTTAGACTTCT 33395  
QY 241 GT 242  
DB 33396 GT 33397

RESULT 3  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL1)  
DEFINITION mRNA, complete cds.

ACCESSION AF127036  
MIM 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS  
SOURCE  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
AUTHORS Agnel, M., Vermet, T. and Culouscou, J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel, M. and Culouscou, J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des Carrières, Rueil-Malmaison 92500, France  
FEATURES  
source  
1..2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue="small intestine; colon"  
1..2826  
/gene="CACCL1"  
5..2749  
/gene="CACCL1"  
/note="bovine epithelial chloride channel homolog"  
/product="calcium-activated chloride channel protein 1"  
/codon\_start=1  
/protein\_id="A025487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPKSSVFLILHLLEGALSNLSLIQNNNGYEGIVAVDPNPV  
EDETLQIKDMVTOASLYLFPAETGKRRFYFNAILIPETKTRADYRPLKTEYMAH  
DIVIAESPPNDDEPYTEMONGCEGRIPILFETLPGKRLAEYGPGRAPVHEMAH  
LRMGVFDENDEKFLYLSNGRIQAVRCAGITGIVYKCGGSCYTRKCTFNKTYGL  
YKGECEFLQSRQTEKASIMFAQHDSTLVECTQDNHKEAPNOMONCINRSIWEVI  
RDSDEFKKTPPTPTOPNPFTSLQIGRIQVCLVDKSGSATGNRLNRLNQAQFL  
LQVLEISVNGKTYLELDNGAGADAKDPGVSRFTYTDINGRSVKRALGCVN  
AARRVYPOOGALYIPGMIENDEIOMNPPEIINKDVOHKQVCFSTSGSGFEVAS  
DVPNAPRPDLPFGQITDLKAEIHGSGSLINLTWAPAGDGDYHGAHKYIIRISIID  
LRDNFESLQVNTTALITKEANSEVEPEFRKENTFENGDTLFLAIQAVKVDLKSEI  
SNIAVSLFIPQIPETPSPDETSAPEPNIHINSTITPGIHILKIMRWIGELDLSTIA  
"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 88.2%; Score 213.4; DB 42; Length 2826;  
Best Local Similarity 97.3%; Pred. No. 4e-34;

Matches 217; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GTTTATCTCCACAGACCTCCGCGACAGACACTACTCTGATGAACGCTGCTCCTTG 60  
DB 2596 GTTTATCTCCACAGACCTCCGCGACAGACACTACTCTGATGAACGCTGCTCCTTG 2655  
QY 61 TCCTAATATTCATATCAACAGACCACTTCCTGCAATTTAAATTTATGTGAA 120  
DB 2656 TCCTAATATTCATATCAACAGACCACTTCCTGCAATTTAAATTTATGTGAA 2715  
QY 121 GTGATAGAGAACTGCACTGTCAATAGCCTAGGCGTGAATTTTGTGAGATTAATA 180  
DB 2716 GTGATAGAGAACTGCACTGTCAATAGCCTAGGCGTGAATTTTGTGAGATTAATA 2775  
QY 181 ATAAATCATTCATCTTTTGTGATTAATAATTTCTTAATTTGATTTAGACTTCT 223  
DB 2776 ATAAATCATTCATCTTTTGTGATTAATAATTTCTTAATTTGATTTAGACTTCT 2818

	RESULT	4	195746	878 bp	DNA	PAT	17-JUL-1998
	LOCUS		195746				
	DEFINITION		Sequence 8 from patent US 5733748.				
	ACCESSION		195746				
	NID		g9340216				
	VERSION		195746..1	GI:3940216			
	KEYWORDS		.				
	SOURCE		. Unknown.				
	ORGANISM		Unknown.				
	REFERENCE		Unclassified.				
	AUTHORS		1 (bases 1 to 878)				
	TITLE		Yu.G. and Rosen,C.				
	JOURNAL		Colon specific genes and proteins				
	FEATURES		Patent: US 5733748-A 8 31-MAR-1998;				
	source		location/Qualifiers				
			1..878				
	BASE COUNT		257 a 179 c 188 g 241 t	13 others			
	ORIGIN						
	Query Match		68.9%;	Score 166.8; DB 5;	Length 878;		
	Best Local Similarity		91.6%;	Pred. No. 9,4e-25;			
	Matches 196:	Conservative	1;	Mismatches 15;	Indels 2;	Gaps 2;	
Oy	1	GTATTTCCTCCAGACTCGGCAAGAGACCCTAAGTCTGAATGAAACGTCGCTCCTTG	60				
Db	625	GTTATATCTCCACAGACTGCCGAGAGACACTAATCTGTATGAAAGCTGCCTCCTTG	684				
Oy	61	T-CCTTAATATTTATATCATCACAGCACCATTCTGGCATTTACATTTTTAAAAATTATGTGGA	119				
Db	665	TGCCATAATTTATATATCAACAACACACCATTCCTGGCATTCACATTTTAAAAATTATGTGGA	744				
Oy	120	AGTGGATAGAGAACAATGCTGAGCTGTCAAATAGCCTAGAGGCGTAATTTTTGTCAGATAATAAA	179				
Db	745	ACTGGTAGGAGAACATCGCATGTGTCATATGNCNTAGGGGTGAATTTTGTGGCGATGAAT-A	803				
Oy	180	AATAATCATTCATCCTTTTTTTTGATTATATAAA	213				
		:					
Db	804	AATAATSATTTCANCCCTTTTTTTGTGTTATAAAA	837				
RESULT	5						
ABO17156							
LOCUS		ABO17156	2937 bp	mRNA	ROD	04-MAR-1999	
DEFINITION		Mus musculus gob-5 mRNA, complete cds.					
ACCESSION		ABO17156					
NID		g3721911					
VERSION		ABO17156.1	GI:3721911				
KEYWORDS		GDB-5.					
SOURCE		Mus musculus adult intestine goblet cell cDNA to mRNA.					
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
		Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.					
REFERENCE		1 (sites)					
AUTHORS		Komiya,T., Tanigawa,Y. and Hirohashi,S.					
TITLE		Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice					
JOURNAL		Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)					
MEDLINE		99160866					
REFERENCE		2 (bases 1 to 2937)					
AUTHORS		Komiya,T., Tanigawa,Y. and Hirohashi,S.					
TITLE		Direct Submission					
JOURNAL		Submitted (26-AUG-1998) to the DDBJ/EMBL/Genbank databases. Tohtsu					
		Komiya, ERIU, JST, Hirohashi Cell Configuration Project; 5-9-9,					
		ToKodai, Tsunuba, Ibakaki 300-2635, Japan					
		(E-mail:t.komehcc@jst.go.jp, Tel:81-298-47-7563,					
		Fax:81-298-47-5226)					
FEATURES							
SOURCE		Location/Qualifiers					
		1..2937					

```

gene
  CDS
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /cell_type="globlet cell"
    /dev_stage="adult"
    /tissue_type="intestine"
    15..2756
    /gene="gob-5"
    15..2756
    /gene="gob-5"
    /codon_start=1
    /protein_id="BAA33743.1"
    /db_xref="PID:d1034712"
    /db_xref="PID:g3721912"
    /db_xref="GI:3721912"
    translation="MESLSPKPYFLLILHLEGLVESLILQNNNGEYIVLIDHDV
EDEALIOHKMDVQASPLYLEFATGKRFRFKVALLIDPSMKAREVPRKLETENK
DIVYTSPLIGNDPEYTHHGACGSKGRIRLHTPPFLAGKRLTQVGPDRITVEHMA
FRMGVFNRYNNDKFTYLSKGRPAVCSAIIIGKNQVRCCQSGSITNGKCVIDRVT
LYKDCVFPDPHONKESIMEINOMINSVECTERNNOENPNQONCINLSTWE
IQOEDQSGKQTPMTAPAPAFYLSQIQIRIVCLVDKSGMLNDRLRMNQASRL
LLOETDEQSWGMVWFDSAAVQSELKOLNSGADBDLILKHLPTVSAGTSTCSGLR
AFYIKRKYIDGSEIYLLDGEDNTISSCFPLVQSGAILHTVLAGRAAELEBOL
KMTGGLQTVSSDOVQNNGLVDAFALSNGMALNQSILQESRYLQNNQMSNV
YDSASGRTFLILITHTPTPTITFDPSGVQNGILDTTIVATLQVPGIKGFWF
ASVIALSSQTLULITYTSRAASATLPITVYVNNKTGFSPSVIVYASIRQASPLI
ASVIALSSVNGKIVTLELLDNGADATKNDGVSRFTAFDANGRSYKIMAGG
TSDRQRAAPKRNAMYIDGWIEDGEVRNMPRPETSYVDQKOLCSFTSSGGSFYAT
VPSAARIPDLPPCOITDLKASIOGONLVNLTWAPGDYDGHGRASNTIRKSTSIIV
LNDHNTSLQVNTTGLIRKEASSEIIFEEELGNTFGNGTDIFIIQAVDSNKSSE
SNIRASVFIFPAOEPRIEDSTPPCDDINSTIIPGIIHVKIMKWLDEMOTYTLGLH
BASE COUNT      860 a      718 c      693 g      666 t
ORIGIN
Query Match      38.0%; Score 92; DB 12; Length 2937;
Best Local Similarity 67.1%; Pred. No. 6.6e-10;
Matches 149; Conservative 0; Mismatches 65; Indels 8; Gaps 1
QY 21 CCCCAGACACACTGCTCTGATGAAAGCTGCTGCTTCTCTTAATATTCATATCACACA 80
Db 2617 CTCAGAGAGCGCCCATTCCTCCCGAAGACTCAACCCCTCTGCTCATACACAGATCACACA 2676
QY 81 GCACCATTCCTGGCATTCACATTTTAAAAAATATGTGGAAGTGGATGGAGAACTGCACG 140
Db 2677 GCACCATTCCTGGCATTCACACGTCGTGGAAGATATGTGGAAGTGGATGGAGAAATGCACG 2736
QY 141 TGTCAATAGCTAGGCGCTGAATTTTGTGCAGATAAATAAATAATCAATTCATCCCTTTT 200
Db 2737 TGACACTGAGGTTTGACATGCAATTTTCAGGCAAGAATACACA-----GTCAATTCCT 2788
QY 201 TTGATTAATAAATTTCTAAAGATATTTTAAAGATTCCTCTGT 242
Db 2789 TTCACTGGAGAAATTTTCTAAAGATATTCATTCAGCTCTCTGT 2830

RESULT 6
AC005139 256172 bp DNA HTG 01-APR-1999
LOCUS AC005139
DEFINITION Plasmodium falciparum chromosome 12, WORKING DRAFT SEQUENCE, 5
unorderded pieces.
ACCESSION AC005139
NID 94558581
VERSION AC005139.3 GI:4558581
KEYWORDS HTG: HTGS PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum.
REFERENCE 1 (bases 1 to 256172)
AUTHORS Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 256172)
AUTHORS Hyman,R.W., Qiu,F., Fung,E.L., Conway,A.B. and Davis,R.W.

```



On Apr 2, 1999 this sequence version replaced gl:4371267  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces

On Mar 24 1999 this sequence version replaced gi:4455715. For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciiparum](http://www.sanger.ac.uk/Projects/P_falciiparum). IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be







OKAIIIVIMADIVAVSFIGLEVGSEFFLEESIKECEYALVIAKELALMSYL  
NISMKNILIPDGIKGRHSHSPMLFOPHYVRLDRHTLKLYMTWQVVRPVST  
LMAIALQICFYPSMSLTFTTIVNSVSALSLVIFYFYFAKELAPHPPLAKFLCIT  
GLTFVFWOGLADILVAMGFIKSHHFWLEVDIOAIONVLVCLEMYIFAANOKRAY  
HAGPYSGETKKIKDKTE"  
complement(3139..3228)  
/gene="F17L22.30"  
/number=1  
complement(3229..3519)  
/gene="F17L22.30"  
/number=2  
complement(3520..3605)  
/gene="F17L22.30"  
/number=2  
complement(3606..3728)  
/gene="F17L22.30"  
/number=3  
complement(3729..3811)  
/gene="F17L22.30"  
/number=3  
complement(3812..4078)  
/gene="F17L22.30"  
/number=4  
7631..9323  
/gene="F17L22.40"  
join(7631..7882,8239..8418,8610..8691,8770..9036,  
9127..9323)  
/gene="F17L22.40"  
/note="similarity to P1g3 Homo sapiens, PID:g2754812;  
contains EST gb:T04375, T20767, AA712527, AA712507"  
/codon\_start=1  
/product="putative NADPH quinone oxidoreductase"  
/protein\_id="CAB36802.1"  
/db\_xref="PID:e1388530"  
/db\_xref="GI:4455266"  
/translation="MKAIIVSEPGKPEVQLRDVADPEYKDEVLIRVATLALNRADT  
LORGLYNPQSSPYLIGLGGTIESVGKSRMVGQVOCALLSGGAVKRVSPA  
GQIFIPRAGISLKDAAAPPEVACTWSTYPMGRSLSVGSLIHGSSSICIPAIOIA  
KHLGVVEFTAGSDERLACKELGADVCIINTEDVAVKAEYTDGKGVLDICGA  
PLQKRLDNLNDFGRLCITGLMGANAELKLSLPRKLTIVGALALRPSPNKAVV  
REVERNVPALIAAGKVPVYKYLPLPSQAEGSHLMESSNHIGKILLET"  
7631..7882  
/gene="F17L22.40"  
/number=1  
7883..8238  
/gene="F17L22.40"  
/number=1  
8239..8418  
/gene="F17L22.40"  
/number=2  
8419..8609  
/gene="F17L22.40"  
/number=2  
8610..8691  
/gene="F17L22.40"  
/number=3  
8692..8769  
/gene="F17L22.40"  
/number=3  
8770..9036  
/gene="F17L22.40"  
/number=4  
9037..9126  
/gene="F17L22.40"  
/number=4  
9127..9323  
/gene="F17L22.40"  
/number=5  
join(10492..10537,10626..10681,10755..11015,11644..11694,  
12714..12894,13180..13235,13404..13487,13641..13733,  
13827..13894,14073..14146,14265..14383)  
/gene="F17L22.50"

/note="similarity to bifunctional nuclease, Zmnia  
elegans, U90266"  
/codon\_start=1  
/product="putative bifunctional nuclease"  
/protein\_id="CAB36803.1"  
/db\_xref="PID:e1388531"  
/db\_xref="PID:g4455267"  
/db\_xref="GI:4455267"  
/translation="MOLMSASENSDTIYHYNITLALMLSLHIGDIHOPLAYGLDE  
GGNTITVRYRRKTNLHHVSVCYRMKKEVIPPDMINSYDLPMMKVDNMIIESALK  
TYNKSPLPMELQANLMTISLGYPLMRDLKRSYEEDTVVAKLLEPSANGE  
LAAVCSMPDEIKKLPQMRWTSALHFPADPDYKCYNEYSNLTALMLSLHYGDIHOP  
LKEGFGIDGKXKIKVHWYNOETNLRHWDDMTISALETTYNSSLPRIHLOAKIK  
NKMNDPSPESCQINQTRCPRPYASESIDLACKYAIRNATAGTTLGDIYFYSRLPYV  
EKRLAOGRLRLGTLNRLIFSARKKLARA  
10492..10537  
/gene="F17L22.50"  
/number=1  
10492..14383  
/gene="F17L22.50"  
10538..10625  
/gene="F17L22.50"  
/number=1  
10626..10681  
/gene="F17L22.50"  
/number=2  
10682..10754  
/gene="F17L22.50"  
/number=2  
10755..11015  
/gene="F17L22.50"  
/number=3  
11016..11643  
/gene="F17L22.50"  
/number=3  
11644..11694  
/gene="F17L22.50"  
/number=4

Query Match 16.9%; Score 40.8; DB 7; Length 107702;  
Best Local Similarity 51.7%; Pred. No. 6.5;  
Matches 93; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY	24	CAGAGACACCTAGCTGATGAAACGTCCTGCTCTGCTCAATATTCATATCAACAGCA	83
DB	56838	CAGATGATCTAAGGATTCAGATATCTCTCCAGAACTAACCCCTTTGTTATAAGC	56897
QY	84	CCATTCCTGGCATTACATTTTAAATATATGGAAGTGAAGAGAGAGACAGCTGCTG	143
DB	56898	CCATTATGATATGCCCCATGTACAAAAGTAAGAGGCCCAACAAGACTTGCATTTGT	56957
QY	144	CAATAGCCTAGGCGCTGAATTTTGTGAGATAAATATAATCAATCATTCCTTTTGT	203
DB	56958	AAATACTATGTTGTTAAATTATATACGCAATATATATATTTCTTGTCTTTTGTG	57017

RESULT	12
TIJ1/c	TIJ1
LOCUS	53778 bp DNA
DEFINITION	Arabidopsis thaliana BAC TIJ1.
ACCESSION	AF128393
NID	G4325340
VERSION	AF128393.1 GI:4325340
KEYWORDS	
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.	
REFERENCE	1 (bases 1 to 53778)
AUTHORS	Washington University Genome Sequencing Center.
TITLE	The A. thaliana Genome Sequencing Project

JOURNAL Unpublished (1997)  
REFERENCE 2 (bases 1 to 53778)  
AUTHORS Scott, K., Bauer, C. and Sandberg, B.  
TITLE The sequence of A. thaliana T1J1  
JOURNAL Unpublished (1999)  
REFERENCE 3 (bases 1 to 53778)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (13-FEB-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63108, USA  
e-mail: rwaterston@wustl.edu

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is T4B21, 200 bp overlap; 3' clone is T32N4. Actual start of this clone is at base position 58066 of T4B21; actual end is at 53778 of T1J1.

Base position 85555 to 85579 of the T1J1 submitted sequence is represented by PCR product from T1J1 genomic DNA only.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFINDER (P. Green and L. Hillier, ms in preparation).

FEATURES  
SOURCE location/Qualifiers  
1..53778  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/clone="T1J1"  
/chromosome="IV"  
/map="unknown"  
complement(1038..4143)  
/gene="T1J1.3"  
complement(join(1038..1439,1526..1621,1749..2020,2075..2303,2398..2601,2696..2797,2884..3324,3402..3590,3711..3873,3923..4143))  
/gene="T1J1.3"  
/note="contains similarity to homeobox domains (Pfam: PF00046, Score,36.5, E=6.9e-08, N=1)"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="A017342.1"  
/db\_xref="PID:g4325343"  
/db\_xref="GI:4325343"  
/translation="MESHMFWTPTKSTNDNDIGTGRREDFTKSGTEVTTENDS  
GEEIDPSQPNKKRRYHRHQIOLESLKFTSYLSTLEFRFKECPHPDCKOR  
KEISRDNLPLQVFWONKRTOMKAQSEHENOILKSDNKLRAENNRKELASNA  
TCPNGGPAIAGSEFDEQHLRIENARLREIDRISIAAKYVGKPGSSFAPLAIA

PSRLDLGVNFGNQTFVEEMTGTGDLISVSIPSETRPIVELAAAMEELVRNA  
QTGDPMLSTDSNVEILNEEYFRTPEPGIKGLRSEASQSAVINHNLVET  
LMDVONSCVSGIVSRALTEVSTGAGVGNALQVMAEFQVPLPTRENTYV  
RYCKHSDGSAVAVDLSLRPSTPLIRRRSRGGLIOELPGVSKYGMIEEMED  
DRSVHNNKPLVOSGLAFGAKRWATPEORCECLASSMASIIPEDLSGRGSDQFKV  
GAKLNEVMTSPSGKSMKLARMTKSTSCGASTAHMTWTSTGSDVAVKTR  
KSMDDPGRPEGVLSAATSEFIVAPRRVDFLDENSRKEMDILSGWQEMAHIA  
NHGEPGCVSLIRVNSGNSQSNMLIOESCSTGASGVYKAPVDIYAMNVLSGDP  
DYVALPLSGFALIPDGSGVGGDGNQHOEMSTSSCGCSILTVARQIIVDSVPTAK  
LSLGSVATVNSLICTVERIKAAVSCVGGGA"  
complement(13764..19734)  
/gene="T1J1.4"  
complement(join(13764..13913,13955..14077,14300..14422,  
14536..14623,14730..14788,14868..14927,15012..15072,  
15158..15289,15570..15780,15998..16131,16333..16415,  
16488..16553,16774..16829,16889..17015,17372..17480,  
17565..17719,17899..18069,18858..18908,18996..19077,  
19158..19241,19294..19340,19588..19734))  
/gene="T1J1.4"  
/note="similar to N-ethylmaleimide sensitive fusion  
proteins; contains similarity to Arpases (Pfam: PF00004,  
Score=307.7, E=1.4e-88n N=1); coded for by A. thaliana  
cDNA N6739"  
/codon\_start=1  
/protein\_id="A017345.1"  
/db\_xref="PID:g4325346"  
/db\_xref="GI:4325346"  
/translation="MAGRYGQVMTMTVTPSADLAFTNLAYCSSSGLRQFVPSGD  
LFLANKRLMSVYMLFIYRGNALNALIORHARVSTGMAVSRRVPEPNEFL  
AMLELEFKKTSKSVDAALSTOLKRENSFLOVTOCKAFEEHGTNYLTVAR  
ADVEGQHTNGIENGILSKPTIYVPEASNGIKIVQRAASNITKKEFNIESG  
IGLGAERFIENRFAFRVPPHVSRLIKHKMLGLPGPTGTLNARQIKML  
NGADPKIVNGPEVLSKFEVGETERNRDLFADADODOTLKEFLFLICASELHVI  
FDEIDAIKSGSRDGTGVNDSIVNOLTKIDGVEALNNVLIGFTNRDLDEALL  
RPGLEVOVEISLPDEAGRLIOIHTNKKKSESLTDNLQELARTEYGAELE  
GVKASATVALNQLMSMDLLKPYEENIKITMEDPLHAYEOPARAGASTDLEER  
LNGVDCGHNNHLYKRAMLIVQKYSTSPVLTCLLEPSSGKRALATGIGDSB  
PFRYKYSATMTGLSESTICAHIVKFEADVSPMSIILDDIERLEITAGPRS  
NIISQTLMLVTLKRPGRKRLVGFTEVTFLESVGISDCEFTVSHSPLOEDAK  
VLNQLNEFSDVDVSAEALNDVSRFLRAEIDIKNREIVLFMPKIKYMLIEMAAGE  
NGSAAEIVAGREKININHFYDCLGDIRFTG"  
21999..28868  
/gene="T1J1.2"  
join(21999..22336,22414..22482,22573..22661,22745..22857,  
22869..23074,23520..23605,23655..23814,23864..24239,  
24282..24389,24598..25216,25263..25390,25714..25937,  
26236..26472,26825..26996,27315..27859,28111..28287,  
28428..28868)  
/gene="T1J1.2"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="A017341.1"  
/db\_xref="PID:g4325342"  
/db\_xref="GI:4325342"  
/translation="MNOQNPREEVSIVNNGGGGITEAPAVEREKEEGLOKOEETI  
ESTDPLIVYEERLIEKSDGEKEDNSSSNNEIDPVSPATFCYKLRQNSLAK  
MSVPELCRNFSAVAWCKLNALACASETCARIPSSKANPFWPIHLLPERTECAV  
ENYVADSPDSVOFIEMSPTCPRALIANFERITIMPTGOSANLVADATSWCE  
HENRODIAYVTKWLTGSAVPGCGPSPQVILGOSQSRFRSRAKAQYALNPNYV  
SESDIOIYVRSIOHVLTLVHPPLVYVILILHLOIGGPTSRMSNGSTPKPFS  
TKKGLGAGSGITMADAITTDSGAMHVAGVPIVNSTIYVMTVGTGPGNGLOATKTI  
STGSRVPSLSSSTGFPRLAAYLRSWDEYILSEIKOGKRPDSDSILKAAQ  
SAATTWGSVTAVADPTRGGSVIAVVIYEGQYMSPPDEPSPITGMVQWESSV  
QPVLAQIFQNPISNFGQVPTQVWVSRYDMSIIPTKDFKNQVAAAGSVDAPEP  
DSGDERKANVYEDPDLPSDIRLARIYVSAHGGEJAIFLRGCVHIFSGPTFSPEN  
YQVNGSAIAPAPSPSCSASVHMDAKDKAMKTIIRVLPALRNSKYNOSIYWM  
RAIERNIKCDLARESAFLNPRMSLIVCMVMDAGVGTQSAAREGTSINSYIYAM  
DADPHSLPSQHQOQYQVCFYSTRQLIIFLFKCLNLCRGCTIYVFAKIIYIYI  
FFWFEVNNDRINCRLLBGMNAGEVRAVMDQARLLDMLGIGESALVNSALVSE  
PFWVDEETITGIPENMAVDPALVSIQAVYAVDLASHFIRLRKRVASFCTLASH  
ASAGSGALNNNTSPTONASSPATQVPPDSILAVOQPTTTTITATNSGSSH  
VQAMGOGALAKISSNDGNSNSTASPSGSPTEPBIINGTGPCTAVRILIGCHFL  
RCQGLLFCLOSSSRFPQRMNDVSSQKLOTGATSLAEVNSAKPFPALNRIEDAGF

```

      gene
      CDS
      /gene="t1j1.1"
      /join(33304..33483,34168..34986)
      /note="similar to the Drosophila DES-1 protein
      (GB:X94180)"
      /codon_start=1
      /evidence=not_experimental
      /protein_id="AADI7340.1"
      /db_xref="PID:g4325341"
      /translation="MGKGRKISSNEEREGVMAIDFMSYTDPEHNRRLIISCY
      /translaltion="MGKGRKISSNEEREGVMAIDFMSYTDPEHNRRLIISCY
      HELSHNASTPTVNRKLGIFANLPRIQVPMSTPOKHEHRRFGQSDIDMDVPTIT
      EHALVTNIAKTIWFLQLEFYALRPIFKPAPGIEWEINFLIQVLDSVYLFEGW
      RSEFALISTFVGGMHPMAGHISEHYVFNPNQETSYSGPLNLTWVGYNHEHD
      PPIPGNKHLHKEIAGEYEGLSESKYSQVITYIMDTVGPYSRMKRKLSKD"
      complement(35280..41120)
      /gene="t1j1.5"
      complement(101n(35280..35429,35540..35754,35848..36073,
      36168..36432,36529..36823,36951..37035,37157..37369,
      37426..37467,37649..37788,38150..38192,38387..38452,
      38925..38997,39275..39357,39443..39618,40001..40244,
      40353..40556,40650..40742,40856..40912,40995..41120))
      /gene="t1j1.5"
      /note="similar to beta-transducins (Pfam: PF00400,
      Score=71.7, E=1.5e-17, N=6)"
      /codon_start=1
      /evidence=not_experimental
      /protein_id="AADI7343.1"
      /db_xref="PID:g4325344"
      /db_xref="GI:4325344"
      /translation="MGTFEPFAIGITSTVPFSVORLCTETFFVVSNGKAQITNCA
      KNLVITISKFFLLKPKLPRKIRALASYRDTYFAFGNIAVFRRAHOVAWSKRVAK
      Query Match
      Best Local Similarity 60.4%; Pred. No. 7.8;
      Matches 67; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
      QY 121 GTGATAGAGAACTGACCTGCAATAGCGCTGAATTTTGTGATAAATAA 180
      I I I I I I I I I I I I I I I I I I I I I I I I I I I I
      Db 52907 GAGCATGTGAAAACCTTCATGATGAGAAAGATGATGATATATATATGCGAA 52848
      QY 181 ATAAATCATTCATCCTTTTGTGATTAAATTTCTAAATGATATTT 231
      I I I I I I I I I I I I I I I I I I I I I I I I I I I I
      Db 52847 ATACATCATTCCTACATTTTCTTAATATTAAATTCATAT 52797
      RESULT 13
      LOCUS ATAC005693 80891 bp DNA PLN 28-OCT-1998
      DEFINITION Arabidopsis thaliana chromosome II BAC T25N22 genomic sequence,
      complete sequence.
      ACCESSION AC005693
      NID G3805755
      VERSION AC005693.1 GI:3805755
      KEYWORDS
      HTG.
      SOURCE
      ORGANISM
      Arabidopsis thaliana
      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
      euhylophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
      eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
      Arabidopsids.
      REFERENCE
      1 (bases 1 to 80891)
      AUTHORS
      Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M.,
      Shen,M., Rensing,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
      TITLE
      Arabidopsis thaliana chromosome II BAC T25N22 genomic sequence
      JOURNAL
      Unpublished
  
```

```

      REFERENCE
      AUTHORS
      TITLE
      JOURNAL
      COMMENT
      2 (bases 1 to 80891)
      Rounsley,S.D. and Lin,X.
      Direct Submission
      Submitted (23-SEP-1998) The Institute for Genomic Research, 9712
      Medical Center Dr., Rockville, MD 20850, USA, rounsley@tigr.org
      3 (bases 1 to 80891)
      Rounsley,S.D.
      Direct Submission
      Submitted (28-OCT-1998) The Institute for Genomic Research, 9712
      Medical Center Dr., Rockville, MD 20850, USA
      On Oct 28, 1998 this sequence version replaced gi:3738273.
      Address all correspondence to:
      Steve Rounsley
      The Institute for Genomic Research
      9712 Medical Center Dr,
      Rockville, MD 20850,
      USA
      e-mail: rounsley@tigr.org
      BAC clone T25N22 is from Arabidopsis chromosome II and is near the
      molecular marker m1310.
      The orientation of the sequence is from SP6 to T7 end of the BAC
      clone.
      Genes were identified by a combination of three methods: Gene
      prediction programs including GRAFT (available by anonymous ftp
      from arthur.epm.cornl.gov), GeneFinder (Phil Green, University of
      Washington), GenScan (Chris Burge,
      http://genomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
      (http://www.cds.dtu.dk/netgene/cdsnetgene.html), searches of the
      complete sequence against a peptide database and the Arabidopsis
      EST database at TIGR (http://www.tigr.org/tdb/at.html).
      Annotated genes are named to indicate the level of evidence for
      their annotation. Genes with similarity to other proteins are named
      after the database hits. Genes without significant peptide
      similarity but with EST similarity are named as 'unknown' proteins.
      Genes without protein or EST similarity, that are predicted by more
      than two gene prediction programs over most of their length are
      annotated as 'hypothetical' proteins. Genes encoding tRNAs are
      predicted by tRNAscan-SE (Sean Eddy,
      http://genome.msu.tl.edu/eddy/tRNAscan-SE/). Simple repeats are
      identified by RepeatMasker (Arjan Smit,
      http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
      genomic sequence that are not annotated as genes but have predicted
      exons by GRAFT are annotated as misc features.
      FEATURES
      source
      1..80891
      /organism="Arabidopsis thaliana"
      /cultivar="Columbia"
      /db_xref="taxon:3702"
      /chromosome="II"
      /map="m1310"
      /clone="T25N22"
      /clone(155..326,583..771,952..1065,1419..1541,1626..1748,
      1823..1921,2066..2122,2241..2287,2282..2525,2635..2812,
      2896..2976,3102..3185,3268..3615)
      /gene="T25N22.1"
      153..3615
      /gene="T25N22.1"
      /note="identical to PIR|A49398"
      /join(165..326,583..771,952..1065,1419..1541,1626..1748,
      1823..1921,2066..2122,2241..2287,2282..2525,2635..2812,
      2896..2976,3102..3185,3268..3387)
      /gene="T25N22.1"
      /codon_start=1
      /product="cycloartenol synthase"
      /protein_id="AAC69112.1"
      /db_xref="PID:g3805756"
      /db_xref="GI:3805756"
      /translation="MACHRCNRYVLPMSGLYGRKRYGPTTSVLSIRKELFVPHDEVN
      WNEANRCAKEDLYYPHPVLVODILWASLHKIVPEVLMWPANLREKIRAIKIHRY
      EDEENRYKICPEPVNKLMLCWEVDEPSEAFKLPRIHDFLWLAEGMGOGYNS
      OLMPDGEIQIALITNLVEEYGPVLEKHSFKNQVLEDCPGDLNRYRHSIGAMP
      FSTDHGPISDCTAEGIKAAALLSKVKAIVGEPIDAKRXYEAVNVIISLONDGGL
      ATYELRISYPMLELINPMTFQDIVIDVPIVECTSAIALISRKLYPGRRKKEVDE
  
```





JOURNAL  
Submitted (23-MAY-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT  
SUBMITTED BY: WOGSC  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone; and the assembly was  
confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
http://www.nhgri.nih.gov/DIR/CHB/CHR7, send  
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by  
Pieter de Jong and coworkers at Roswell Park Cancer Institute,  
using the method described by Ioannou et al., Nature Genetics  
6:84-9 (1994). The library is from one male donor. For further  
details, see http://bacpac.med.buffalo.edu/  
The clone is available from Genome Systems, Inc.  
(http://www.genomesystems.com).  
VECTOR: pCYPAC2

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is D00547C10. Actual start of this  
clone is at base position 1 of D0807C15; actual end is at 11752  
of D0807C15.

This clone contains STS SWS1919 (NID:g1113280).

This clone contains a transposon, which has been cut from the  
submitted sequence, from base 83390 to base 84661.

#### FEATURES

##### source

```
1. 117752
   location/Qualifiers
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /map="7q34-q36"
   /clone_lib="RPCI-4"
   /clone="D0807C15"
   1. 1663
   /rpt_family="L1"
   repeat_region
   1664. 1963
   /rpt_family="Alu"
   repeat_region
   1964. 2042
   /rpt_family="L1"
   repeat_region
   2049. 2262
   /rpt_family="L1"
   repeat_region
   2295. 2591
   /rpt_family="Alu"
   repeat_region
   2592. 2987
   /rpt_family="L1"
   repeat_region
   2984. 3712
   /rpt_family="L1"
   repeat_region
   3724. 3973
```

```
repeat_region
4436. 4815
/rpt_family="MALR"
repeat_region
5001. 5974
/rpt_family="L1"
repeat_region
6210. 6502
/rpt_family="Alu"
repeat_region
8817. 10017
/rpt_family="MER2_type"
repeat_region
10668. 10943
/rpt_family="MER1_type"
repeat_region
13463. 13582
/rpt_family="Alu"
repeat_region
15402. 15858
/rpt_family="L1"
repeat_region
15843. 15900
/rpt_family="L1"
repeat_region
15901. 15952
/rpt_family="Retroviral"
repeat_region
16052. 16439
/rpt_family="Retroviral"
repeat_region
16441. 16533
/rpt_family="L1"
repeat_region
16545. 16559
/rpt_family="L1"
repeat_region
16691. 17037
/rpt_family="MALR"
repeat_region
17040. 17322
/rpt_family="Alu"
repeat_region
17271. 17362
/rpt_family="Alu"
repeat_region
17366. 17405
/rpt_family="MALR"
repeat_region
17406. 19021
/rpt_family="MALR"
repeat_region
19022. 19419
/rpt_family="MALR"
repeat_region
19439. 19781
/rpt_family="L1"
misc_feature
21253. 21423
/rpt_family="MER1_type"
repeat_region
22015. 22319
/rpt_family="Alu"
repeat_region
23330. 23639
/rpt_family="Alu"
repeat_region
25095. 25352
/rpt_family="Alu"
repeat_region
25579. 25747
/rpt_family="MIR"
repeat_region
26687. 26972
/rpt_family="Alu"
repeat_region
27102. 27464
/rpt_family="Retroviral"
repeat_region
27465. 28103
/rpt_family="Retroviral"
repeat_region
28324. 28623
/rpt_family="Alu"
repeat_region
30520. 30637
/rpt_family="L1"
repeat_region
30893. 30948
/rpt_family="L1"
repeat_region
31377. 31653
/rpt_family="Alu"
repeat_region
32851. 33309
/rpt_family="MALR"
repeat_region
33323. 33621
/rpt_family="Alu"
repeat_region
33983. 34037
/rpt_family="L2"
repeat_region
34907. 35075
/rpt_family="Alu"
```



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:04:31 ; Search time 921.96 Seconds  
(without alignments)  
65,671 Million cell updates/sec

Title: US-09-049-696-16

Perfect score: 242  
Sequence: 1 GTTATTCTCCACAGACTC.....ATGATTTCAGACTTCCTCTGT 242

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201.2	83.1	255	1 T22483	Human gene signatu
2	166.8	68.9	878	1 T45884	Human colon specif
3	166.8	68.9	878	1 V16772	Polynucleotide seq
4	36.8	15.2	110000	1 V21209_00	Methanococcus jann
5	34.8	14.4	110000	1 X20248_08	Continuation (9 of
6	34.4	14.2	134525	1 O04525	Total base sequenc
7	34.4	14.2	110000	1 X20248_05	Continuation (6 of
8	34.4	14.2	116277	1 X20249	Borrelia burgdorfe
9	34	14.0	110000	1 X20248_00	Borrelia burgdorfe
10	33.8	14.0	805	1 O83839	B. burgdorferi T5
11	33.4	13.8	7074	1 X20266	Borrelia burgdorfe
12	33.2	13.7	2520	1 O77884	Neural thread prot
13	33.2	13.7	2520	1 T27766	AD 16c human neuro
14	33.2	13.7	10684	1 T33758	Control region iso
15	33.2	13.7	110000	1 V21209_04	Continuation (5 of
16	33.2	13.7	110000	1 X20248_03	Continuation (4 of
17	32.8	13.6	6217	1 V65252	DNA encoding S. pn
18	32.6	13.5	19124	1 T72882	Plasmodium var-7 g
19	32.2	13.3	1452	1 X20315	Borrelia burgdorfe
20	32.2	13.3	110000	1 X20248_06	Continuation (7 of
21	32.2	13.3	111309	1 X20250	Borrelia burgdorfe
22	32	13.2	110000	1 V21209_07	Continuation (8 of
23	32	13.2	53585	1 X20251	Borrelia burgdorfe
24	31.8	13.1	773	1 N71219	Sequence of the 5'
25	31.8	13.1	773	1 N71220	Sequence of the 5'
26	31.8	13.1	773	1 N70212	Sequence of 5' flia
27	31.8	13.1	773	1 N70211	Hepatitis A virus
28	31.8	13.1	7494	1 V05898	P. vulgaris chondr
29	31.6	13.1	6519	1 O74990	Proteus vulgaris c
30	31.6	13.1	6519	1 T58982	DNA encoding a fer
31	31.6	12.8	1298	1 T83779	DNA encoding a Sta
32	31	12.8	638	1 T83780	Continuation (3 of
33	31	12.8	110000	1 V21209_02	Continuation (3 of
34	31	12.8	631	1 V75003	Staphylococcus aur
35	30.8	12.7	110000	1 V21209_05	Continuation (6 of
36	30.8	12.7	64976	1 V21209_16	Continuation (17 of
37	30.6	12.6	110000	1 T58840_1	Continuation (2 of
38	30.6	12.6	110000	1 T58840_3	Continuation (4 of
39	30.6	12.6	110000	1 V21209_13	Continuation (14 of
40	30.6	12.6	2129	1 V52849	Human eyal gene co
41	30.6	12.6	5849	1 V31335	Plasmodium berghei
42	30.6	12.6	8155	1 V74374	Staphylococcus aur
43	30.4	12.6	7493	1 O03512	Attenuated hepatit

ALIGNMENTS

44 30.4 12.6 6768 1 028301  
45 30.4 12.6 329 1 059715

AMEPV Spheroidin D  
Human brain Expres

RESULT 1  
ID T22483 standard; cDNA to mRNA; 255 BP.  
AC T22483;  
DE 22-AUG-1996 (first entry)  
DE Human gene signature H0MG504094.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; Cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PI (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR MPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 1138-1139; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 255 BP; 76 A; 52 C; 39 G; 80 T;

Query Match 83.1%; Score 201.2; DB 1; Length 255;  
Best Local Similarity 95.8%; Pred. No. 6.3e-43;  
Matches 203; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
2 TTTATTCCTCCACAGACTCCGACAGACCTAGTCTGATGAACGCTCTCTTGT 61  
|||||  
43 TTTATTCCTCCACAGACTCCGACAGACCTAGTCTGATGAACGCTCTCTTGT 102  
|||||  
62 CCTATATTCATATCAACAGACCACTTCTGCAATTTTAAATTTATGGAAG 121  
|||||  
103 CCTATATTCATATCAACAGACCACTTCTGCAATTTTAAATTTATGGAAG 162  
|||||  
122 TCGATATGGAAGTCAAGTCTGCAATAGCTGAGGCTGAATTTTGTGATTAATAA 181  
|||||  
163 TCGATATGGAAGTCAAGTCTGCAATAGCTGAGGCTGAATTTTGTGATTAATAA 222  
|||||  
182 TAAATCATTCATCTTTTATTTTATTAATAA 213  
|||||  
223 TAAATCATTCATCTTTTATTTTATTAATAA 254  
|||||  
RESULT 2  
T45884  
ID T45884 standard; cDNA; 878 BP.  
AC T45884;

DT 13-MAR-1997. (first entry)  
 DE Human colon specific gene CSG5 cDNA partial clone.  
 KW Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;  
 KM gene therapy; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 2..691  
 FT /\*tag= a  
 PN W0639419-A1.  
 PD 12-DEC-1996.  
 PF 06-JUN-1995; U07289.  
 PR 06-JUN-1995; MO-U07289.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Yu G.  
 DR WPI: 97-043054/04.  
 P-PSDB: W06548.  
 PT Human colon specific genes and their expression products - detection  
 PT of which, in non-colon tissue samples, can be used as indication of  
 PT colon cancer metastasis.  
 PS Claim 1: Fig 5; 60pp; English.  
 CC 13 cDNA clones (743880-92), most of them partial clones, correspond  
 CC to human colon specific genes, designated CSG1, CSG2, etc., that  
 CC are primarily expressed in tissues derived from the colon. CSG7  
 CC and CSG10 show reduced expression in colon cancer cells as compared  
 CC to that in normal cells; the remaining genes are overexpressed in  
 CC colon cancer. The partial cDNA sequences can be used to isolate  
 CC full-length clones and genomic clones including the complete gene.  
 CC CSG nucleic acids can be used to produce CSG polypeptides (see also  
 CC W06545-53) in transformed host cells, as probes to detect disorders  
 CC of the colon, partic. colon cancer and colon cancer metastasis, and  
 CC in gene therapy.  
 SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 68.9%; Score 166.8; DB 1; Length 878;  
 Best Local Similarity 91.6%; Pred. No. 4.6e-34;  
 Matches 196; Conservative 1; Mismatches 15; Indels 2; Gaps 2;

QY 1 GTTTATTCCTCCACAGACTCCGCCGAGACACTAGTCTGATGAACGCTGCTCCTTG 60  
 DB 625 GTTATTCCTCCACAGACTCCGCCGAGACACTAGTCTGATGAACGCTGCTCCTTG 684  
 QY 61 T-CCTAATATTCATATCAACAGACACCATTCCTGCGATTCACATTTTAAATTTATGTGGA 119  
 DB 685 TGCCATAATTCATATCAACAGACACCATTCCTGCGATTCACATTTTAAATTTATGTGGA 744  
 QY 120 AGTGATAGGAGAGACTGCACTGTCATAGCCTAGGCTGAATTTTGTGCAGATAAATA 179  
 DB 745 AGTGGTAGGAGAACTGCAGTGTCAATAGNCTAGGGGTGAATTTTGTGCGGTGAAT-A 803  
 QY 180 AATAATCATTCATCCCTTTTGTGATTAATAA 213  
 DB 804 AATAATCATTCATCCCTTTTGTGATTAATAA 837

RESULT 3  
 V16672  
 ID V16672 standard; cDNA; 878 BP.  
 AC V16672.  
 DT 22-JUN-1998 (first entry)  
 DE Polynucleotide sequence of a colon-specific gene.  
 KW Colon-specific gene; probe; detection; expression; human;  
 KM diagnostic assay; colon cancer; antibody; screening; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 2..685  
 FT /\*tag= a  
 FT /note= "no stop codon given"  
 PN US5733748-A.  
 PD 31-MAR-1998.  
 PF 06-JUN-1995; 469667.  
 PR 06-JUN-1995; US-469667.  
 PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen C, Yu G;  
 DR WPI: 98-229823/20.  
 DR P-PSDB: W46879.  
 PT Colon-specific nucleic acids - useful as probes for detecting colon  
 PT cancer micrometastases  
 PS Claim 15; Fig 5A-B; 51pp; English.  
 CC V16668-81 represent polynucleotide sequences of partial or full-length  
 CC cDNA clones of colon-specific genes. The polynucleotides can be used  
 CC as probes to detect expression of the corresponding human genes, e.g. in  
 CC diagnostic assays for detecting micrometastases of colon cancer.  
 CC Recombinant cells containing the polynucleotides can be used to  
 CC produce the polypeptides, in order that antibodies can be raised and  
 CC used in further screening or diagnostics.  
 SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 68.9%; Score 166.8; DB 1; Length 878;  
 Best Local Similarity 91.6%; Pred. No. 4.6e-34;  
 Matches 196; Conservative 1; Mismatches 15; Indels 2; Gaps 2;

QY 1 GTTTATTCCTCCACAGACTCCGCCGAGACACTAGTCTGATGAACGCTGCTCCTTG 60  
 DB 625 GTTATTCCTCCACAGACTCCGCCGAGACACTAGTCTGATGAACGCTGCTCCTTG 684  
 QY 61 T-CCTAATATTCATATCAACAGACACCATTCCTGCGATTCACATTTTAAATTTATGTGGA 119  
 DB 685 TGCCATAATTCATATCAACAGACACCATTCCTGCGATTCACATTTTAAATTTATGTGGA 744  
 QY 120 AGTGATAGGAGAGACTGCACTGTCATAGCCTAGGCTGAATTTTGTGCAGATAAATA 179  
 DB 745 AGTGGTAGGAGAACTGCAGTGTCAATAGNCTAGGGGTGAATTTTGTGCGGTGAAT-A 803  
 QY 180 AATAATCATTCATCCCTTTTGTGATTAATAA 213  
 DB 804 AATAATCATTCATCCCTTTTGTGATTAATAA 837

RESULT 4  
 V21209\_00/c  
 WP Sequence split into 17 fragments LOCUS V21209 Accession V21209  
 WP Fragment Name Begin End  
 WP V21209\_00 1 110000  
 WP V21209\_01 100001 210000  
 WP V21209\_02 200001 310000  
 WP V21209\_03 300001 410000  
 WP V21209\_04 400001 510000  
 WP V21209\_05 500001 610000  
 WP V21209\_06 600001 710000  
 WP V21209\_07 700001 810000  
 WP V21209\_08 800001 910000  
 WP V21209\_09 900001 1010000  
 WP V21209\_10 1000001 1110000  
 WP V21209\_11 1100001 1210000  
 WP V21209\_12 1200001 1310000  
 WP V21209\_13 1300001 1410000  
 WP V21209\_14 1400001 1510000  
 WP V21209\_15 1500001 1610000  
 WP V21209\_16 1600001 1664976  
 ID V21209 standard; DNA; 1664976 BP.  
 AC V21209.  
 DT 10-NOV-1998 (first entry)  
 DE Methanococcus jannaschii circular chromosome.  
 KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;  
 KM genome; autotrophic; extrachromosomal element; identification; ds.  
 OS Methanococcus jannaschii.  
 PN M09807830-A2.  
 PD 26-FEB-1998.  
 PF 22-AUG-1997; U14900.  
 PR 22-AUG-1996; US-024428.  
 PA (GENO-) INST GENOMIC RES.  
 PA (UNIV ) UNIV ILLINOIS FOUND.  
 PA (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 PI Bilt CJ, Smith HO, Venter JC, White OR, Woese CR;

DR WPI: 98-169145/15  
PT Complete genome sequence of methano-genic archaeon, Methanococcus  
PT jannaschii - useful in identification of M. jannaschii genome  
PT fragment  
PS Claim 13: Page 152-585; 614pp; English.  
CC The present sequence represents the complete 1.66-megabase pair genome  
CC sequence of the Methanococcus jannaschii circular chromosome. The  
CC present invention describes M. jannaschii open reading frames from the  
CC genome sequence. The invention also describes a computer based system  
CC for identifying fragments of the M. jannaschii genome that are  
CC homologous to target nucleotide sequences, comprising: (a) data storage  
CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550  
CC bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at  
CC least 99.9% identical to it; (b) search means for comparing a target  
CC sequence to the nucleotide sequence of the data storage means to  
CC identify a homologous sequence, and (c) retrieval means for obtaining  
CC the homologous sequence. The method, which is based on whole genome  
CC random sequencing of an autotrophic archaeon M. jannaschii, the genome  
CC of which consists of 3 physically distinct elements, a large circular  
CC chromosome (the 1664976 bp sequence given in V21209), a large circular  
CC extra-chromosomal element (the 58407 bp sequence given in V21210), and a  
CC small circular extra-chromosomal element (the 16550 bp sequence given in  
CC V21211), can be used in the identification of M. jannaschii genome  
CC fragment.  
SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T;

Query Match 15.2%; Score 36.8; DB 1; Length 110000;  
Best Local Similarity 55.5%; Pred. No. 1.5;  
Matches 71; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 112 TATGTGGAAGTGGATGGAGAACTGCAGCTGTCATAGCCTAGGCGTGAATTTTGTACAG 171  
DB 38254 TAT 38195

OY 172 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 231  
DB 38194 AGNCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 38135

OY 232 AGACTTCC 239  
DB 38134 AAACATTC 38127

RESULT 5  
X20248\_08

Continuation (9 of 10) of X20248 from base 800001 (Borrelia burgdorferi polynucleotide  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

WP	Fragment Name	Begin	End
WP	X20248_00	1	110000
WP	X20248_01	100001	210000
WP	X20248_02	200001	310000
WP	X20248_03	300001	410000
WP	X20248_04	400001	510000
WP	X20248_05	500001	610000
WP	X20248_06	600001	710000
WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

Query Match 14.4%; Score 34.8; DB 1; Length 110000;  
Best Local Similarity 57.3%; Pred. No. 4.9;  
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

OY 91 TGGCATTTCACATTTTAAATTAATGAGAGTGGATGAGAGAACTGCAGCTGCATFAC 150  
DB 22712 TATGTACAAATATATATATCTTTTTCATATAGGGGTTTTCATATTCATTGGC 22771

OY 151 CTAGGCGTGAATTTTGTGAGATAAATAAATATCATTCATCTTTT 200  
DB 22772 GCATGTTTGAATTTTTCAGACAACTAAAGAAAGCCCATATTCATT 22821

RESULT 6  
Q04525  
AC Q04525 standard; DNA; 134525 BP.  
ID Q04525;  
DT 01-OCT-1990 (first entry)  
DE Top base sequence of rice plant chloroplast DNA.  
KW Chloroplast; true grass; rice plant; ss.  
OS Oryza sativa.  
PN J02100682-A.  
PF 12-APR-1990.  
PR 07-OCT-1988; 251967.  
PA 07-OCT-1988; JP-251967.  
PT (MITR) Mitsui Toatsu Chem Inc.  
DR WPI: 90-159709/21.  
PT Chloroplast DNA of true grasses - used to produce various  
PT DNA base sequences by decomposition of rice plant DNA.  
PS Claim 1: Fig 1: 20pp; Japanese.  
CC The sequence is that of the whole of rice chloroplast DNA.  
SQ Sequence 134525 BP; 41249 A; 26129 C; 26331 G; 40816 T;

Query Match 14.2%; Score 34.4; DB 1; Length 134525;  
Best Local Similarity 57.4%; Pred. No. 6.5;  
Matches 62; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 100 CATTTTAAATTAATGAGAGTGGATGAGAGAACTGCAGCTGTCATATAGCCTAGGCGTG 159  
DB 44132 CCTATTTCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 44191

OY 160 AATTTTGTGACATAAATAAATTAATCATTCATCTTTTTCATT 207  
DB 44192 CTTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 44239

RESULT 7  
X20248\_05

Continuation (6 of 10) of X20248 from base 500001 (Borrelia burgdorferi polynucleotide  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

WP	Fragment Name	Begin	End
WP	X20248_00	1	110000
WP	X20248_01	100001	210000
WP	X20248_02	200001	310000
WP	X20248_03	300001	410000
WP	X20248_04	400001	510000
WP	X20248_05	500001	610000
WP	X20248_06	600001	710000
WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

Query Match 14.2%; Score 34.4; DB 1; Length 110000;  
Best Local Similarity 50.6%; Pred. No. 6.3;  
Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY 67 TATTCATATCAACAGCACCATTCCTGCGATTCACATTTTAAATTAATGATGAGAGTGAT 126  
DB 36428 TTTTCATATATATATTAAGACCTCTGTTTCATTTTACATTTTATGTTTAAAGTGCT 36487

OY 127 AGGAGAACTGCGAGCTGTCATATAGCCTAGGCGTGAATTTTGTCAATATAATTAAT 186  
DB 36488 ACATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 36547

OY 187 CATTCATCTTTTGTGATTAATAAATTTCTAATAATGATTT 230  
DB 36548 GGGAGAAATTTTAT 36591

RESULT 8  
X20249/C  
ID X20249 standard; DNA; 116277 BP.  
AC X20249;

DE 04-MAY-1999 (first entry)  
 DE Borrelia burgdorferi polynucleotide sequence #2.  
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.  
 OS Borrelia burgdorferi.  
 PN M09858943-A1.  
 PD 30-DEC-1998.  
 PF 18-JUN-1998; U12764.  
 PR 03-SEP-1997; US-057483.  
 PR 20-JUN-1997; US-050354.  
 PR 22-JUL-1997; US-053344.  
 PR 22-JUL-1997; US-053377.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,  
 PI White OR;  
 PI WPI: 99-081217/07.  
 DR New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease  
 PS Claim 1; Page 672-737; 1128pp; English.  
 CC X20248 to X20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.  
 SO Sequence 116277 BP; 42656 A; 19868 C; 14430 G; 39250 T;

Query Match 14.2%; Score 34.4; DB 1; Length 116277;  
 Best Local Similarity 50.6%; Pred. No. 6.3;  
 Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY 67 TATTCATTCACACGACCATTCCTCGCATTCACATTTTAAATAATGCGAAGTGCAT 126  
 DB 18639 TTTTCATTTTATTAAGACCTCTGTTTCAATTTTAACTTTTAAAGTGGT 18580  
 OY 127 AGGAGACGACGAGCTGCTAATAGCTAGGCGTGAATTTTGCACATAATAATAAT 186  
 DB 18579 ACAAATAATTTATTTATTTGTTAACTTTTAACTTTTAACTTTTAACTTTTAA 18520  
 OY 187 CATTCATCCCTTTTGTGATTAATAAATTTCTAAATGTAATT 230  
 DB 18519 GCGAGAAATTTTATGTATATAAAATGTTTTTAAACATAATT 18476

RESULT 9  
 X20248-00  
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
 WP Fragment Name Begin End  
 WP X20248\_00 1 110000  
 WP X20248\_01 100001 210000  
 WP X20248\_02 200001 310000  
 WP X20248\_03 300001 410000  
 WP X20248\_04 400001 510000  
 WP X20248\_05 500001 610000  
 WP X20248\_06 600001 710000  
 WP X20248\_07 700001 810000  
 WP X20248\_08 800001 910000  
 WP X20248\_09 900001 910715  
 ID X20248 standard; DNA; 910715 BP.  
 AC X20248;  
 DE 04-MAY-1999 (first entry)  
 DE Borrelia burgdorferi polynucleotide sequence #1.  
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.  
 OS Borrelia burgdorferi.

PN M09858943-A1.  
 PD 30-DEC-1998.  
 PF 18-JUN-1998; U12764.  
 PR 03-SEP-1997; US-057483.  
 PR 20-JUN-1997; US-050359.  
 PR 22-JUL-1997; US-053344.  
 PR 22-JUL-1997; US-053377.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,  
 PI White OR;  
 PI WPI: 99-081217/07.  
 DR New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease  
 PS Claim 1; Page 157-671; 1128pp; English.  
 CC X20248 to X20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.  
 SO Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T;

Query Match 14.0%; Score 34; DB 1; Length 110000;  
 Best Local Similarity 50.6%; Pred. No. 7.9;  
 Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

OY 72 ATATTCACACGACCATTCCTCGCATTCACATTTTAAATAATGCGAAGTGCAT 131  
 DB 26415 ATATTATCACCAGAAATAGCTGCAAAATCAAAATAAATTAATGATTTAGAAA 26474  
 OY 132 AACGCGACGCTGCTAATAGCCGAGCGTGAATTTTGCACATAATAATAATCATT 191  
 DB 26475 AAGAAAAATATATATGAATTAAGAGCAATTTATTAATGTTTAAAAAAGTAGAAA 26534  
 OY 192 ATCCCTTTTGTGATTAATAAATTTCTAAATGTAATTAG 233  
 DB 26535 GCGAAATTTTCTTAATAGAGAAAAAATTAATAATTTTG 26576

RESULT 10  
 X20248-00  
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
 WP Fragment Name Begin End  
 WP X20248\_00 1 110000  
 WP X20248\_01 100001 210000  
 WP X20248\_02 200001 310000  
 WP X20248\_03 300001 410000  
 WP X20248\_04 400001 510000  
 WP X20248\_05 500001 610000  
 WP X20248\_06 600001 710000  
 WP X20248\_07 700001 810000  
 WP X20248\_08 800001 910000  
 WP X20248\_09 900001 910715  
 ID X20248 standard; DNA; 805 BP.  
 AC 083839;  
 DE 05-SEP-1995 (first entry)  
 DE B. burgdorferi T5 gene.  
 KW Lyme disease; borreliosis; immunization; vaccine; diagnostic; ds.  
 OS Borrelia burgdorferi N40.  
 FH Key Location/Qualifiers  
 FT cds 120..711  
 FT /\*tag- a  
 PN M09504145-A.  
 PD 09-FEB-1995.  
 PR 26-JUL-1994; U08529.  
 PR 30-JUL-1993; US-099757.  
 PR 08-SEP-1993; US-118469.  
 PA (UYVA) UNIV VALE.  
 PI Barthold SW, Fikrig E, Flavell RA, Kantor FS, Lam TT;  
 DR WPI: 95-082229/11.  
 DR P-PDB; R69631.  
 PT New DNA encoding new Borrelia burgdorferi polypeptide(s) - and  
 PT related vectors, transformed cells, fusion or multimeric proteins  
 PT and antibodies, for treatment, prevention and diagnosis of Lyme  
 PT disease  
 PS Claim 1; Page 84-85; 114pp; English.  
 CC A cDNA library in lambda ZAPit was constructed from B. burgdorferi  
 CC N40 and screened with rabbit anti-serum against N40. 6 positive  
 CC clones, not contg. OspA, OspB or flagellin genes, were selected.

CC 1 of these contained the r5 gene (Q88833) that encoded a 21.8 kDa  
CC protein (R65633). The protein may be used in the development of  
CC multimeric vaccines.

50 Sequence	805 BP;	266 A;	96 C;	120 G;	303 T;
-------------	---------	--------	-------	--------	--------

**SQ** Sequence 805 BP; 286 A; 96 C; 120 G; 303 T;

Query Match	14.08;	Score 33.8;	DB 1;	Length 805;
Best Local Similarity	49.78;	Pred. No. 3.1;		
Matches	86;	Conservative	0;	Mismatches 87;
			Indels	0;
			Gaps	0;

[illegible]

	RESULT	11
ID	X20266	/c
	X20266/c	
AC	X20266	standard; DNA; 7074 BP.
DE	04-May-1999	(first entry)
DT	Borrelia burgdorferi	polynucleotide sequence #19.
DR	Borrelia burgdorferi	spirochetes; bacterium; pathogen; Lyme disease
KM	epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis	
KN	infection; diagnosis; characterization; detection; ds.	

Query Match	13.8%;	Score 33.4;	DB 1;	Length 7074;
Best Local Similarity	62.7%;	Pred. No. 6.2;		
Matches 52;	Conservative 0;	Mismatches 31;	Indels 0;	Gaps 0;

Accession	Sequence	Position
Oy	TGAATTTTTCACATTAATAAATAATCATCCCTTTTCTTTTGATTAATAAATTTT	217
158	TT	
Db	TGATTAATAATATACCTTTTAGTTAAATAAATTTTATTCACATTTAACTTTT	5277
Oy	CTAAAATGTATTAGACTTCCT	240
218	TT	
Db	TTAAATTTTATTTATATATCTC	5254

RESULT	12
Q77884/c	
ID	Q77884 standard; cDNA; 2520 BP.

[illegible]

## RESULT 13

T27766/c  
ID T27766 standard; cDNA; 2520 BP

CC tumours and a malignant astrocytoma in a human. A number of clones  
CC of neural thread protein were isolated from healthy 17-18 week old  
CC foetal human brain (HB) 2 year old temporal lobe neocortex and endo  
CC stage Alzheimer's disease (AD) cerebral cortex. See 1271753-75.  
SQ Sequence 2520 BP; 661 A; 577 C; 581 G; 701 T;

Db 9339 TATATTATTATATGTTGATCAACCCATGTT 9306  
RESULT 15  
V21209\_04

Query Match	13.7%	Score 33.2;	DB 1;	Length 2520;
Best Local Similarity	57.8%	Pred. No. 5.6;		
Matches	59;	Conservative	0;	Mismatches 43;
			Indels	0;
			Gaps	0;

Fragment Name	Begin	End	LOCUS	Accession	Fragment Size
Fragment 01	100001	210000	V21209	V21209	109999
Fragment 00	100001	110000	V21209	V21209	9999
Fragment 17	100001	110000	V21209	V21209	9999
Fragment 15	100001	110000	V21209	V21209	9999
Fragment 14	100001	110000	V21209	V21209	9999
Fragment 13	100001	110000	V21209	V21209	9999
Fragment 12	100001	110000	V21209	V21209	9999
Fragment 11	100001	110000	V21209	V21209	9999
Fragment 10	100001	110000	V21209	V21209	9999
Fragment 09	100001	110000	V21209	V21209	9999
Fragment 08	100001	110000	V21209	V21209	9999
Fragment 07	100001	110000	V21209	V21209	9999
Fragment 06	100001	110000	V21209	V21209	9999
Fragment 05	100001	110000	V21209	V21209	9999
Fragment 04	100001	110000	V21209	V21209	9999
Fragment 03	100001	110000	V21209	V21209	9999
Fragment 02	100001	110000	V21209	V21209	9999
Fragment 01	100001	110000	V21209	V21209	9999
Fragment 00	100001	110000	V21209	V21209	9999
Fragment 17	100001	110000	V21209	V21209	9999
Fragment 15	100001	110000	V21209	V21209	9999
Fragment 14	100001	110000	V21209	V21209	9999
Fragment 13	100001	110000	V21209	V21209	9999
Fragment 12	100001	110000	V21209	V21209	9999
Fragment 11	100001	110000	V21209	V21209	9999
Fragment 10	100001	110000	V21209	V21209	9999
Fragment 09	100001	110000	V21209	V21209	9999
Fragment 08	100001	110000	V21209	V21209	9999
Fragment 07	100001	110000	V21209	V21209	9999
Fragment 06	100001	110000	V21209	V21209	9999
Fragment 05	100001	110000	V21209	V21209	9999
Fragment 04	100001	110000	V21209	V21209	9999
Fragment 03	100001	110000	V21209	V21209	9999
Fragment 02	100001	110000	V21209	V21209	9999
Fragment 01	100001	110000	V21209	V21209	9999
Fragment 00	100001	110000	V21209	V21209	9999

Qy	153	Agggctgaattttgtcagataaaataaaataatcattatc	194
Db	2352	AAAAAGAAATCTGAAATGAATTAATTAACAAGAAAGCTGATGAT	2311
Qy	93	GCAATTCACATTTTAAAAATATATGCGAAGTGATAGAGAAACGCGAGCTCAAAATAGCT	152
Db	2412	GAAACATGCTTTAAATATATCCAGTGGAGGAGAGGGGGTGAATTAACACGTCAACAAAGCTA	2355

RESULT	14
ID	T33758/c
AC	T33758 standard; DNA; 10684 BP
DT	19-DEC-1996 (first entry)

Query Match	13.7%	Score	33.2	DB	1	Length	110000
Best Local Similarity	49.4%	Pred.	No. 13				
Matches	86	Conservative	0	Mismatches	88	Indels	0
						Gaps	0

PE 19-MAR-1996: U03808.  
PR 20-MAR-1995: US-408584.  
PR 05-APR-1995: US-418096.  
PR 02-AUG-1995: US-510584.  
PR 30-OCT-1995: US-558588.  
PR 21-NOV-1995: US-007390.  
PR 30-NOV-1995: US-007721.  
PR 14-DEC-1995: US-008601.  
PR (INSP) INST PASTEUR LILLE.  
PA (LIGA-) LIGAND PHARM INC.  
PI Auerx J, Briggs MR, Croston GE, De Vos P, Miller SG;

QY	59	TGTCCTAAATTCATCTTCACAGCCACATCTCGCATCCATTAAATTAATGTGG	118
Db	61647	TTTCAGAAAGTGGAAATTCAGATATTATTATATTTCGACCCAAACATTCCAAAGTATAAGG	61700
QY	119	AACTGGATAGGAGAACTGCAGCTGCAATAGCCCTAGGCGCTGAAATTTTGTGCAGATAATA	178
Db	61707	AATTTGAAATTAAGACATCAATCAATGTCAGAGTTTATAGATATCCAGCATTTGGAATTAATTC	61766
QY	179	AAATTAATCATTCACCTCTTTTGTGATTATAAAATTTTCTAAATGTATTTTA	232
Db	61767	CAATATTCTCAGTCCAAATATTTTTCAATATAAAATTTTGGAGAGATGTTTTTTA	61820

Search completed: August 6, 1999, 04:05:20  
Job time: 1672 sec

Query Match	13.7%	Score	33.2	DB	1	Length	10684
Best Local Similarity	59.6%	Pred. No.	7.7				
Matches	56	Conservative	0	Mismatches	38	Indels	0
						Gaps	0

Qy 138 AGCTGTCAATAGCCTAGGGCTGAATTTTGTGCAGATAAATAAATAATCAATCATCCTT 197  
 11 11 1111 11 1111 11 11 111111 11 11 11  
 Db 9399 AGATGGCAATATTAGTAGCAATTTCATTTCTGGCAATACAAATAAAAACAAGCAGCTGTGTCTA 9340



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:13:14; Search time 921.96 Seconds

(without alignments)  
809.496 Million cell updates/sec

Title: US-09-049-696-20

Perfect score: 2983

Sequence: 1 GAAATCAGGAGGAGTGTAC.....AAATGCTAACACTGGGTA 2983

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database: N.Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	790.8	26.5	878	1 T45884	Human colon specif
2	790.8	26.5	878	1 V16672	Polynucleotide seq
3	241.2	8.1	255	1 T22483	Human gene signatu
4	70	2.3	398	1 X40511	Human secreted pro
5	40.2	1.3	5156	1 V10238	Arabidopsis ferula
6	38.5	1.3	5156	1 T90505	Arabidopsis thalia
7	37.8	1.3	110000	1 V21209_07	Continuation (8 of
8	37.8	1.3	110000	1 X20248_03	Continuation (4 of
9	36.8	1.2	1052	1 N90224	Continuation (4 of
10	36.8	1.2	7320	1 T68081	Malattia-specific D
11	36.8	1.2	110000	1 V21209_00	H. pylori cytoplas
12	36.5	1.2	2854	1 V53366	Methanococcus jann
13	36.6	1.2	2424	1 V74703	DNA encoding a Sta
14	36.4	1.2	8700	1 Q42541	Stephylococcus aur
15	36.4	1.2	5849	1 V31135	BgIII/HpaII fragme
16	36	1.2	3158	1 Q31880	Plasmodium berghei
17	36	1.2	3158	1 O53212	Cyclin D3 promoter
18	36	1.2	110000	1 X20248_05	Human cyclin D3 pr
19	36	1.2	116277	1 X20249	Continuation (6 of
20	35.8	1.2	1671	1 Q24134	Borrelia burgdorfe
21	35.4	1.2	4505	1 V15448	50 kD subunit of S
22	35.2	1.2	64976	1 V21209_16	Arabidopsis PR-1 p
23	35.2	1.2	2822	1 X52222	Continuation (17 o
24	35	1.2	1794	1 T33873	Protein PRO228 cDN
25	34.8	1.2	6924	1 O67283	Candida RAP1 cDNA
26	34.8	1.2	805	1 O67283	Human RPR-beta. H
27	34.8	1.2	19124	1 T72882	B. burgdorferi 15
28	34.8	1.2	110000	1 X20248_00	Plasmodium var-7 g
29	34.8	1.2	110000	1 X20248_08	Borrelia burgdorfe
30	34.4	1.2	134525	1 O04525	Continuation (9 of
31	34.4	1.2	7074	1 X20266	Total base sequenc
32	34.2	1.1	2736	1 O84047	Borrelia burgdorfe
33	34.2	1.1	877	1 V75289	Insertion sequence
34	34	1.1	5319	1 V35235	Staphylococcus aur
35	33.8	1.1	1566	1 O84697	R. prowazekii S-la
36	33.8	1.1	2255	1 O84698	Human IFN-gamma ac
37	33.8	1.1	3447	1 T08711	Human IFN-gamma ac
38	33.6	1.1	1200	1 O52267	Sak-a serine-threo
39	33.6	1.1	4090	1 O85027	Sequence encoding
40	33.6	1.1	3183	1 T42919	Entamoeba histolyt
41	33.6	1.1	5391	1 V71729	SAG12-1 promoter,
42	33.6	1.1	5407	1 V71738	Upstream sequence
43	33.6	1.1	4810	1 V70887	Plasmid pTK-fus-1a

44 33.6 1.1 1177 1 X20333 Borrelia burgdorfe  
c 45 33.6 1.1 110000 1 X20248\_03 Continuation (4 of

## ALIGNMENTS

```
RESULT 1
ID T45884
AC T45884;
DE 13-MAR-1997 (first entry)
DE Human colon specific gene CSG5 cDNA partial clone.
KM Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;
KW gene therapy; ss.
OS Homo sapiens.
FH Key
FT cds Location/Qualifiers
FT 2..691
FT /*tag= a
FN WO9639419-A1.
PD 12-DEC-1996.
PE 06-JUN-1995; U07289.
PR 06-JUN-1995; WO-U07289.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Yu G.
DR WPI: 97-043054/04.
DR P-PSDB: W06548.
PT Human colon specific genes and their expression products - detection
PT of which, in non-colon tissue samples, can be used as indication of
PT colon cancer metastasis
PS Claim 1: Fig 5; 60pp; English.
CC 13 cDNA clones (T45880-92), most of them partial clones, correspond
CC to human colon specific genes, designated CSG1, CSG2, etc., that
CC are primarily expressed in tissues derived from the colon. CSG7
CC and CSG10 show reduced expression in colon cancer cells as compared
CC to that in normal cells; the remaining genes are overexpressed in
CC colon cancer. The partial cDNA sequences can be used to isolate
CC full-length clones and genomic clones including the complete gene.
CC CSG nucleic acids can be used to produce CSG polypeptides (see also
CC W06545-53) in transformed host cells, as probes to detect disorders
CC of the colon, partic. colon cancer and colon cancer metastasis, and
CC in gene therapy.
SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match
Best Local Similarity 97.9%; Pred. No. 6.4e-211;
Matches 820; Conservative 1; Mismatches 15; Indels 2; Gaps 2;
26.5%; Score 790.8; DB 1; Length 878;

OY 1992 TGCTACCTCAAGGATTTTACAACTTATGACACGATGGTAGATACAGTGTAAAGTCG 2051
DB 1 TGCTACCTCAAGGATTTTACAACTTATGACACGATGGTAGATACAGTGTAAAGTCG 60
OY 2052 GGCTGGGAGAGGATTAACGCGACGACGAGGAGTGTATACCCGACGAGGAGGACAT 2111
DB 61 GGCTGGGAGAGGATTAACGCGACGACGAGGAGTGTATACCCGACGAGGAGGACT 120
OY 2112 GTACATACCTGGCTGATGATGAGATGATGATGATGATGATGATGATGATGATGAT 2171
DB 121 GTACATACCTGGCTGATGATGAGATGATGATGATGATGATGATGATGATGATGAT 180
OY 2172 TAATTAAGATGATGTTTCAACACACGAAAGTGTTCACGACACATCTCGGAGGCTC 2231
DB 181 TAATTAAGATGATGTTTCAACACACGAAAGTGTTCACGACACATCTCGGAGGCTC 240
OY 2232 ATTTGGCTTGTGATGTCACCAATGCTCCATACCTGATCTCTCCACCTGGCCAAAT 2291
DB 241 ATTTGGCTTGTGATGTCACCAATGCTCCATACCTGATCTCTCCACCTGGCCAAAT 300
OY 2292 CACGACCTGAAGGCGGAATTCACGCGGCGAGTCTCATATATCTGACACTGGACAGCTCC 2351
DB 301 CACGACCTGAAGGCGGAATTCACGCGGCGAGTCTCATATATCTGACACTGGACAGCTCC 360
```

```

OY 2352 TGGGATGATATGACCATGGAAGCTCACAAGATATCATTCGAATTAAGTACAGTAT 2411
    |||||
Db 361 TGGGATGATATGACCATGGAAGCTCACAAGATATCATTCGAATTAAGTACAGTAT 420
OY 2412 TCTTATCTCAGAGCAAGATTCATGAATCTCTTCAAGGATAGTACTAGCTCTCATCCC 2471
    |||||
Db 421 TCTTATCTCAGAGCAAGATTCATGAATCTCTTCAAGGATAGTACTAGCTCTCATCCC 480
OY 2472 AAAGGAGCCAACTCTGAGGAAGCTTTTGTAAACCCAGAAAACATTACTTTGAAAA 2531
    |||||
Db 481 AAAGGAGCCAACTCTGAGGAAGCTTTTGTAAACCCAGAAAACATTACTTTGAAAA 540
OY 2532 TGGCAGACATCTTTCATTCGATTCAGGCTGTGTATAGTCGATCTGAAATCGAAAT 2591
    |||||
Db 541 TGGCAGACATCTTTCATTCGATTCAGGCTGTGTATAGTCGATCTGAAATCGAAAT 600
OY 2592 ATCCACATTCGACAGATATCTTTTATTCCTCCACAGACTCCGCGAGAGACACTAG 2651
    |||||
Db 601 ATCCACATTCGACAGATATCTTTTATTCCTCCACAGACTCCGCGAGAGACACTAG 660
OY 2652 TCCGATGAAACGTCCTCTCTGT-CCTAATATTCATATCAACAGACACATTCCTGCA 2710
    |||||
Db 661 TCCGATGAAACGTCCTCTCTGT-CCTAATATTCATATCAACAGACACATTCCTGCA 720
OY 2711 TTCACATTTTAAATATGAGGATGAGATGAGAACTGCGATGCAATAGCCTAG 2770
    |||||
Db 721 TTCACATTTTAAATATGAGGATGAGATGAGAACTGCGATGCAATAGCCTAG 780
OY 2771 GCTGAATTTTGTGAGATTAATTAATATCATTCATCTTTTGTGATTAATAA 2828
    |||||
Db 781 GCTGAATTTTGTGAGATTAATTAATATCATTCATCTTTTGTGATTAATAA 837

RESULT 2
V16672 2
ID V16672 standard; cDNA; 878 BP.
AC V16672:
DE 22-JUN-1998 (first entry)
DE Polynucleotide sequence of a colon-specific gene.
KW Colon-specific gene; probe; detection; expression; human;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 2..685
FT tag-a
FT /note="no stop codon given"
PN US5733748-A.
PD 31-MAR-1998.
PE 06-JUN-1995: 469667.
PR 06-JUN-1995: US-469667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen C, Xu G;
DR WPI: 98-229823/20.
DR P-PSDB: W46879.
PT Colon-specific nucleic acids - useful as probes for detecting colon
PT cancer metastases
PS Claim 15: Fig 5A-B: 51pp; English.
CC V16668-81 represent polynucleotide sequences of partial or full-length
CC cDNA clones of colon-specific genes. The polynucleotides can be used
CC as probes to detect expression of the corresponding human genes, e.g. in
CC diagnostic assays for detecting the micrometastases of colon cancer.
CC Recombinant cells containing the polynucleotides can be used to
CC produce the polypeptides, in order that antibodies can be raised and
CC used in further screening or diagnostics.
SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 26.5%; Score 790.8; DB 1; Length 878;
Best Local Similarity 97.9%; Pred. No. 6.4e-211;
Matches 820; Conservative 1; Mismatches 15; Indels 2; Gaps 2;
OY 1992 TGTCTACTCAAGGATATTTCACAACTTATGACACGAATGATAGATCAAGTAAAGTGGC 2051
    |||||

```

```

Db 1 TGTCTACTCAAGGATATTTCACAACTTATGACACGAATGATAGATCAAGTAAAGTGGC 60
OY 2052 GGCCTCGGAGAGATTAACGACGACAGAGAGATATACCCAGAGATGGAGCACT 2111
    |||||
Db 61 GGCCTCGGAGAGATTAACGACGACAGAGAGATATACCCAGAGATGGAGCACT 120
OY 2112 GTACATACCTGGCTGATGGAATGATGAATTAATTAATTAATTAATTAATTAATTAAT 2171
    |||||
Db 121 GTACATACCTGGCTGATGGAATGATGAATTAATTAATTAATTAATTAATTAATTAAT 180
OY 2172 TAATAGGATGATGTTAACAACAGCAAGTGTGTTGAGCAAGCAATCTCGGAGGCTC 2231
    |||||
Db 181 TAATAGGATGATGTTAACAACAGCAAGTGTGTTGAGCAAGCAATCTCGGAGGCTC 240
OY 2232 ATTGTGGCTCTGATGTCCCAATGCTCCCATACCTGATCTCTCCACCTGGCCAAAT 2291
    |||||
Db 241 ATTGTGGCTCTGATGTCCCAATGCTCCCATACCTGATCTCTCCACCTGGCCAAAT 300
OY 2292 CACGACCTGAAGCGGAAATTCACGGGGCAGTCTCATTAATCTGACTTGACAGCTCC 2351
    |||||
Db 301 CACGACCTGAAGCGGAAATTCACGGGGCAGTCTCATTAATCTGACTTGACAGCTCC 360
OY 2352 TGGGATGATTAATGACATGGAACAGCTCAACAGTATATCATTCGATTAATGATCAAGTAT 2411
    |||||
Db 361 TGGGATGATTAATGACATGGAACAGCTCAACAGTATATCATTCGATTAATGATCAAGTAT 420
OY 2412 TCTTATCTCAGAGCAAGTCAATGAATCTCTTCAAGTAAATACTACTCTCATCCC 2471
    |||||
Db 421 TCTTATCTCAGAGCAAGTCAATGAATCTCTTCAAGTAAATACTACTCTCATCCC 480
OY 2472 AAAGGAGCCAACTCTGAGGAAGCTTTTGTAAACCCAGAAAACATTACTTTGAAAA 2531
    |||||
Db 481 AAAGGAGCCAACTCTGAGGAAGCTTTTGTAAACCCAGAAAACATTACTTTGAAAA 540
OY 2532 TGGCAGACATCTTTCATTCGATTCAGGCTGTGTATAGTCGATGGAATCAGAAAT 2591
    |||||
Db 541 TGGCAGACATCTTTCATTCGATTCAGGCTGTGTATAGTCGATGGAATCAGAAAT 600
OY 2592 ATCCACATTCGACAGATATCTTTTATTCCTCCACAGACTCCGCGAGAGACACTAG 2651
    |||||
Db 601 ATCCACATTCGACAGATATCTTTTATTCCTCCACAGACTCCGCGAGAGACACTAG 660
OY 2652 TCCGATGAAACGTCCTCTCTGT-CCTAATATTCATATCAACAGACACATTCCTGCA 2710
    |||||
Db 661 TCCGATGAAACGTCCTCTCTGT-CCTAATATTCATATCAACAGACACATTCCTGCA 720
OY 2711 TTCACATTTTAAATATGAGGATGAGATGAGAACTGCGATGCAATAGCCTAG 2770
    |||||
Db 721 TTCACATTTTAAATATGAGGATGAGATGAGAACTGCGATGCAATAGCCTAG 780
OY 2771 GCTGAATTTTGTGAGATTAATTAATATCATTCATCTTTTGTGATTAATAA 2828
    |||||
Db 781 GCTGAATTTTGTGAGATTAATTAATATCATTCATCTTTTGTGATTAATAA 837

RESULT 3
T22483 3
ID T22483 standard; cDNA to mRNA; 255 BP.
AC T22483:
DE 22-AUG-1996 (first entry)
DE Human gene signature HMG504094.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PE 11-NOV-1994: J01916.
PR 12-NOV-1993: JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUBO/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.

```

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 1138-1139; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in 119001-176837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 235 BP; 76 A; 52 C; 39 G; 80 T;

Query Match  
Best Local Similarity 8.1%; Score 241.2; DB 1; Length 255;  
Matches 243; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2575 GATCTGAAATCAGAAATATCCAACTTGCAGAGATCTTTTATCTCCACAGACT 2634  
DB 1 GATCTGAAATNAGAAATATCCAACTTGCAGAGATCTTTTATCTCCACAGACT 60  
QY 2635 CCGCCAGAGACACCTAGTCTGTATGAAGCTGCTCTTCTTATATATATATCAG 2694  
DB 61 CCGCCAGAGACACCTAGTCTGTATGAAGCTGCTCTTCTTATATATATCAG 120  
QY 2695 AGCACCATTCTGATTCATTTTAAATATATGTAAGTGAATAGAGAACTGCAG 2754  
DB 121 AGCACCATTCTGATTCATTTTAAATATATGTAAGTGAATAGAGAACTGCAG 180  
QY 2755 CTGTCAATAGCCTAGGCTGAAATTTTGTCCAGATAAATAATCAATTCCTTTT 2814  
DB 181 CTGTCAATAGCCTAGGCTGAAATTTTGTCCAGATAAATAATCAATTCCTTTT 240  
QY 2815 TTTTGAATTAATAA 2828  
DB 241 TTTTNGTNTTATATA 254

RESULT 4  
X40511  
ID X40511 standard; cDNA; 398 BP.  
AC X40511;  
DT 18-JUN-1999 (first entry)  
DE Human secreted protein 5' EST SEQ ID No: 111.  
KM Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KM forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
KM upstream regulatory sequence; cytokine activity; cell proliferation;  
KM differentiation; haematopoiesis regulation; tissue growth regulation;  
KM reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KM thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN WO9806550-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1232.  
PR 01-AUG-1997; US-905144.  
PA (GEST) GENSET.  
PI Ducleit A, Dumas Milne Edwards J, Lacroix B;  
DR WPI: 99-153780/13.  
DR P-PSDB: Y111789.  
PT New isolated prostate-derived nucleic acids - used to develop  
PT products which may have cytokine, immune regulatory, haematopoiesis  
PT regulating, anti-inflammatory or tumour inhibition activity  
PS Claim 1: Page 228; 675pp; English.

CC X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins expressed in prostate, and encode the proteins given in  
CC Y11176 to Y11937 respectively. The proteins given represent the signal  
CC peptide and an N-terminal fragment of a secreted protein. The nucleic  
CC acid sequences can be used for producing secreted human gene products.  
CC They can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell proliferation and  
CC differentiation activity, haematopoiesis regulating activity, tissue  
CC growth regulating activity, reproductive hormone regulating activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition  
CC activity or other activities. The products can be used in forensic, gene  
CC therapy and chromosome mapping procedures. The sequences can also be used  
CC for obtaining corresponding promoter sequences. The nucleic acids  
CC encoding the signal peptides can be used for directing extracellular  
CC secretion of a polypeptide or the insertion of a polypeptide into a  
CC membrane, or importing a polypeptide into a cell.  
SQ Sequence 398 BP; 123 A; 86 C; 76 G; 113 T;

Query Match  
Best Local Similarity 2.3%; Score 70; DB 1; Length 398;  
Matches 109; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 98 TTCAGCTGAAACAAATGGCTATGAGGATGTCTTGCATTCGACCCCAATGTCCAG 157  
DB 220 TTCAGCTGAAACAAATGGCTATGAGGATGTCTTGCATTCGACCCCAATGTCCAG 279  
QY 158 AGATGAAACACATTCATCAACAAATGAGACATGTGACCAAGCATCTGTATCTGT 217  
DB 280 AGATGAAACACATTCATCAACAAATGAGACATGTGACCAAGCATCTGTATCTGTAT 339  
QY 218 TTGAGCTACAGAAAGCATTTTATCAAAAATGTTCCATTTGATTCCTG 271  
DB 340 TTATGCTACCAAGAAAGATATTTTTCAGAAATTAAGATTTTAAATACCTG 393

RESULT 5  
V10238/c  
ID V10238 standard; cDNA; 5156 BP.  
AC V10238;  
DT 20-JUL-1998 (first entry)  
DE Arabidopsis ferulate-5-hydroxylase genomic DNA.  
KM Ferulate-5-hydroxylase; cinnamate-4-hydroxylase;  
KM tissue-specific promoter; ligand; syringyl; delignification;  
KM pulp; transgenic plant; ss.  
OS Arabidopsis thaliana.  
FH Key Location/Qualifiers  
FT CDS 2487..4522  
FT FT /\*tag= a  
FT FT /\*note= "contains introns"  
FT FT 2997..3125  
FT FT /\*tag= b  
FT FT /number= 1  
FT FT 3557..3901  
FT FT /\*tag= c  
FT FT /number= 2  
PN WO9803535-A1.  
PD 29-JAN-1998.  
PF 18-JUL-1997; U12624.  
PR 16-DEC-1996; US-032908.  
PR 19-JUL-1996; US-022228.  
PA (PURD) PURDUE RES FOUND.  
PI Chapple CS;  
DR WPI: 98-120689/11.  
DR P-PSDB: W40099.  
PT Tissue-specific regulatory plant promoter linked to  
PT ferulate-5-hydroxylase gene - useful for manipulation of plant  
PT lignin composition  
PS Claim 14; Page 48-51; 71pp; English.  
CC This genomic DNA sequence includes a coding region, interrupted  
CC by 2 introns, that codes for the ferulate-5-hydroxylase (F5H, see  
CC W40099) of Arabidopsis thaliana. It comprises the HindIII-XhoI



	Query Match	Similarity	1.3%	Score 37.8;	DB 1;	Length 110000;
	Best Local Similarity		48.4%;	Pred. No. 10;	Mismatches 105;	Conservative 0; Indels 0; Gaps 0;
Oy	403	GAGAAAGGTTGAAGAATCCACCTCACCCTCGATTTCATTCAGAAAAAAGTTAGCTGAA	462			
Dg	66736	GAAGAAGGGTGAAAATAGTGTGTCTTAATGTGATGTGTATCAAGAAAATAATATTGCAT	66795			
Oy	463	TATGACCAACAAGTAGGGCATTTGTCATGAGTGGGCTCATCTACGATGGGAGATATT	522			
Dg	66796	AATTGACCAAGATGGAGGGCATTTCACCATGAGCAAAAGATTGAATGTAGACTTGA	66855			
Oy	523	GACGAGTCAAATATGATGAGAAATTTCTATTCATCCAATGAGAAGATCAACAGCTAAGA	582			
Dg	66856	GCTCTATGACTTATAGTGTCTTATTACACAGAACCAATATATTAAAGACATGAGAA	66915			
Oy	583	TGTTGAGCAGGTATTACTGTGTACAAATGTAGTAAGA	619			
Dg	66916	ATAAAGTTGTATAAATTGGAGACTTATAGATAAA	66952			
	RESULT					
	8					
	X20248_03					
	Continuation (4 of 10) of X20248 from base 300001 (Borrelia burgdorferi polynucleotide s					
WP	Sequence split into 10 fragments	Locus X20248 Accession X20248				
Fragment Name	Begin	End				
WP X20248_00	1	110000				
WP X20248_01		210000				
WP X20248_02		200001	310000			
WP X20248_03		300001	410000			
WP X20248_04		400001	510000			
WP X20248_05		500001	610000			
WP X20248_06		600001	710000			
WP X20248_07		700001	810000			
WP X20248_08		800001	910000			
WP X20248_09		900001	910715			

Query Match	Similarity	1.3%	Score	37.8	DB 1	Length	110000
Best Local	Similarity	45.9%	Pred. No.	10			
Matches	129	Conservative	0	Mismatches	152	Indels	0
						Gaps	0
QY	2679	TAAATATCATATCAACACACCACCATTCCTGGCATTCACATTTTAAATAATATATGGAGATG	2738				
DB	77062	TAAATTTTAAAAAAGAAAAGATTTTCAAAATCTCAAGCTCCATCATATTTTCAAAATTA	77121				
QY	2739	GATPAGAGAACTGCAGCTGTCATATAGCCTAGGGCTGAAATTTTGTGATGATTAATTAATA	2798				
DB	77122	TATATTTTAAAAATAGCATTTGAAAGCTCAAAACCAAGGTTTTCACAAACTTTAACTA	77181				
QY	2799	AATCATATATCTTTTTTTGTGATTTAAATTTTCTTAAATATTTTAGACTTCCTGTA	2858				
DB	77182	TAGGAGCTCCCTGCTCTTTTGTGATTTATCATTTTAAAAATAAAGATTTAAACTTTATCAA	77241				
QY	2859	GGGGGCGATATATCTAAATGATATATAGTACATTTATATCTAAATATTTCTCTGAGGGGCG	2918				
DB	77242	CTATACCTTTTTCAAAATTTAGATTAATTTGAATCTACAGATTCATTTGCCATCAAAATATC	77301				
QY	2919	ATATCTAAATGTATTTTAGACTTCCTGTAAGGGGCGATAA	2959				
DB	77302	TATTTAAATATACATCAAGAACTTCATATTTTGGAAATAA	77342				
RESULT	9						
ID	N90224	standard: DNA: 1052 BP.					
AC	N90224:						
DE	1-NOV-1989	(first entry)					
DT	Malaysia-specific DNA insert of clone 41-2.						
KW	DNA: malaya; clone 41-2; Plasmodium falci-parum; protein: vaccine.						
OS	Plasmodium falci-parum						
PN	EP-322712-A.						

PD	05-JUL-1989 .
PF	20-DEC-1988; 121299.
PR	30-DEC-1987; DE-831351.
RA	(BERM) Behringwerke.
PI	Knaupp B, Hundt E, Enders B, Kupper H;
DR	WPI: 89-194071/27 .
P-PSDB:	P90417 .
PT	New antigenic proteins from Plasmodium falciparum - new encoding nucleic acid sequences and derived antibodies, useful in vaccines, diagnosis etc.
PS	Claim 1; Table 17; 25pp; German.
CC	The DNA is of a malaria-specific DNA insert of clone 41-2. The encoded peptide (see P90417) is useful in a protective vaccine, esp. against malaria. Produced antibodies are useful for passive immunisation, and antibodies, the protein and the DNA sequence are useful as diagnostic reagents. See also N90211-25 and P90403-19.
CC	
CC	
CC	
SQ	Sequence 1052 BP; 394 A; 93 C; 88 G; 477 T;

  

Query Match	1.2%;	Score 36.8;	DB 1;	Length 1052;
Best Local Similarity	50.0%;	Pred. No. 1.7;		
Matches	92;	Conservative	0;	Mismatches 92;
			Indels	0;
			Gaps	0.

  

QY	2710 ATTCGATTTTAAATAATTATGTGAGTGAGGAGACCTGCAGTCATGAGCGTAG	2769
Db	154 ATACATGTTGTGAAGAAGTGTCAAAAATTTATATGAAAAAATAAATTAATGCTAA	213
QY	2770 GGCTGAAATTTTGTGCAGATTAATTAATAATCATTCATCTTTTTTGGATTATAAAT	2829
Db	214 TATATATATATATATATATATATATATATCTTCTTCTTCTTTTTTTTTTTTGT	273
QY	2830 TTCTGAAAATGATTATTAGACTTCCTGTAGGGGGCGATATACTAATGTATAGTACT	2889
Db	274 ATTATTAATGTAFTATTATCTTATGCATGGGATTATTTAACAAATTAATGATAAAT	333
QY	2890 TTAT 2893	
Db	334 AAAT 337	

```

RESULT 10
ID T68081 standard; DNA: 7320 BP.
AC T68081;
DE 16-JUL-1997 (first entry)
KW H. pylori cytoplasmic protein. ltrgl0309orf15.
KW Cytoplasmic; vaccine; prevention; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
OS Helicobacter pylori.
FH Key
FT 1. 7320
FT cds
FT location/Qualifiers
FT /*tag= "a
FT /note= "no stop codon given"
PN WO9640893-A1.
PD 19-DEC-1996.
PE 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI; 97-052306/05.
DR P-PSDB; W20828.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 9; Page 887-889; 1481pp; English.
CC The present sequence encodes a Helicobacter pylori cytoplasmic protein
CC that may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,

```

CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
SQ Sequence 7320 BP; 2828 A; 1317 C; 1436 G; 1739 T;

Query Match 1.2%; Score 36.8; DB 1; Length 7320;  
Best Local Similarity 46.8%; Pred. No. 4.7; Mismatches 132; Indels 0; Gaps 0;  
Matches 116; Conservative 0;

775 ACAGAACAAACACACAAAGAGCTCCAAACAAATCAAAATCAATCCGA 834  
DB 1288 ACCACACAAAGAACCAACACACAAAGAACCAACACACAAAGAACCA 1347  
QY 835 ACCACATGGAGTATCCGTTCTGAGACTTTAAGAAAACCACTCTATGACACA 894  
DB 1348 ACCACACAAAGAACCAACACACAAAGAACCAACACACAAAGAACCA 1407  
QY 895 CAGCCACCAATCCACCTCTCATCTGCTGACAGATTGACAAAGATTGTGTTAGTC 954  
DB 1408 AACACAAAGAACCAATTCAGCAGACATCCCTTAATCTTAATGATTTAACCG 1467  
QY 955 CTGACAAATCTGGAAGCATGCGGACTGTTAACCGCTCAATGCACTGATACAGCAGC 1014  
DB 1468 AGCAGAGAACTTTATATGAGGAGGAGCTAAACTCGCTACAGCTTAAGCATT 1527  
QY 1015 CAGCTTTT 1022  
DB 1528 GAGCTTTT 1535

RESULT 11  
V21209.00/c  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209  
WP Fragment Name Begin End  
WP V21209\_00 1 11000  
WP V21209\_01 100001 210000  
WP V21209\_02 200001 310000  
WP V21209\_03 300001 410000  
WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000  
WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000  
WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976  
WP V21209 standard: DNA; 1664976 BP.  
ID V21209.  
AC V21209.  
DT 10-NOV-1998 (first entry)  
DE Methanococcus jannaschii circular chromosome.  
KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;  
KW genome; autotrophic; extrachromosomal element; identification; ds.  
OS Methanococcus jannaschii.  
PN M09607/830-A2.  
PD 26-FEB-1998.  
PE 22-AUG-1997; U14900.  
PR 22-AUG-1996; US-024428.  
PA (GENO-) INST GENOMIC RES.  
PA (UNIT-) UNIT ILLINOIS FOUND.  
PA (UTJO-) UNIT JOHNS HOPKINS SCHOOL MEDICINE.  
PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;  
DR WPI; 98-169145/15.  
PT Complete genome sequence of methano-genic archaeon, Methanococcus

PT jannaschii - useful in identification of M. jannaschii genome  
PR fragment  
PS Claim 13; Page 152-585; 614pp; English.  
CC The present sequence represents the complete 1.66-megabase pair genome  
CC sequence of the Methanococcus jannaschii circular chromosome. The  
CC present invention describes M. jannaschii open reading frames from the  
CC genome sequence. The invention also describes a computer based system  
CC for identifying fragments of the M. jannaschii genome that are  
CC homologous to target nucleotide sequences, comprising: (a) data storage  
CC means comprising the nucleotide sequences of the 1664976, 58407 or 16550  
CC bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at  
CC least 99.9% identical to it; (b) search means for comparing a target  
CC sequence to the nucleotide sequence of the data storage means to  
CC identify a homologous sequence, and (c) retrieval means for obtaining  
CC the homologous sequence. The method, which is based on whole genome  
CC random sequencing of an autotrophic archaeon M. jannaschii, the genome  
CC of which consists of 3 physically distinct elements, a large circular  
CC chromosome (the 1664976 bp sequence given in V21209), a large circular  
CC extra-chromosomal element (the 58407 bp sequence given in V21210), and a  
CC small circular extra-chromosomal element (the 16550 bp sequence given in  
CC V21211), can be used in the identification of M. jannaschii genome  
SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T;

Query Match 1.2%; Score 36.8; DB 1; Length 110000;  
Best Local Similarity 55.5%; Pred. No. 19;  
Matches 71; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 2727 TATGAGAGTGTAGGAGAGACTGTCATATGCGCTAAGGCGGAATTTTGTGAC 2786  
DB 38254 TATATATGATTTATATATATGATTTATGATTTCTATGAGATTTTGAAGTTATTTAG 38195  
QY 2787 ATAAATATAATTAATCAATCCCTTTTGTATATAAATTTCTAAATGATTTT 2846  
DB 38194 AGACATATAAATAATATAAATAATATATGTTTATTAAGCTCTTCAAAATCTTTT 38135  
QY 2847 AGACTTTC 2854  
DB 38134 AAACATTC 38127

RESULT 12  
V53366  
ID V53366 standard: DNA; 2854 BP.  
AC V53366;  
DT 30-OCT-1998 (first entry)  
DE DNA encoding a Staphylococcus aureus protein of unknown function.  
KW Staphylococcus aureus protein; immune response induction; eye infection;  
KW antibody production; T-cell immune response; gastrointestinal infection;  
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
KW central nervous system; kidney infection; urinary tract infection;  
KW antimicrobial compound identification; broad spectrum antibiotic;  
KW therapy; ss.  
OS Staphylococcus aureus.  
PN EP-841394-A2.  
PD 13-MAY-1998.  
PE 24-SEP-1997; 307485.  
PR 24-SEP-1996; US-027032.  
PA (SMIK) SMITHKLINE BEECHAM CORP.  
PA (SMIK) SMITHKLINE BEECHAM PLC.  
PI Black MT, Burnham MKR, Hodgson JE, Knowles DC,  
PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RM, Rosenberg M,  
PI Ward JM;  
PI WPI; 98-252940/23.  
DR P-PSDB; W75566.  
PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
PT useful in vaccines and for treatment of bacterial infections of e.g.  
PT respiratory tract and central nervous system  
PS Claim 1; Page 69-70; 390pp; English.  
CC This sequence encodes a Staphylococcus aureus protein of unknown  
CC function, and represents a DNA sequence of the invention.  
CC The DNA sequences were isolated from Staphylococcus aureus WCHU29

CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
CC produce polypeptides or fragments. The proteins are used in the treatment  
CC of disease, for inducing an immune response by administering them, to  
CC produce antibody and/or T-cell immune response. Antagonists of the  
CC proteins are used for the inhibition of bacterial polypeptides.  
CC Conditions which may be treated include bacterial infections, especially  
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
CC urinary tract, skin, bones and joints. The proteins can also be used to  
CC identify antimicrobial compounds which are broad spectrum antibiotics,  
CC especially useful in the treatment of H. pylori infection.  
SQ Sequence 2854 BP; 1012 A; 446 C; 453 G; 881 T;

Query Match 1.2%; Score 36.6; DB 1; Length 2854;  
Best Local Similarity 48.9%; Pred. No. 3.2;  
Matches 90; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 2759 CATACCCGAGGCTGATTTTGTGAGATTAATAATTAATCATCTTTT 2818  
DB 1263 CATGCGGAGGAGCTGATTTTGTGACNACTATCAAAATTTCTCANATGCTGTGATTTT 1322  
QY 2819 GATTATAAATTTCTAAATGATTTAGACTTCTGTAGGGGCGATTAATGAT 2878  
DB 1323 TATAANNNGGTAGAGCAAGTATATATCCCTTCAATCGTTTCATATTTATATAT 1382  
QY 2879 ATATAGTACATTTATCTAATGATTTCTGTAGGGGCGATTAATGATTTT 2938  
DB 1383 TTTTAAATATAAGCGCTAATGTTTAACTAAGCATAGATGACAAAGTATATACAG 1442  
QY 2939 ACTT 2942  
DB 1443 AATT 1446

## RESULT 13

V74703  
ID V74703 standard; DNA; 2424 BP.

AC 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #392.  
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
OS Staphylococcus aureus.  
FH Key Location/Qualifiers  
FT misc\_feature 1141..1200

FT /\*tag= a  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence"

PD 30-JUL-1997.  
PF 07-JAN-1997; 100117.  
PR 05-JAN-1996; US-009861.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
PI Rosen CA;  
PI WPI; 97-374922/35.  
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
PT stored on computer readable medium and used in the production of  
PT anti-S. aureus vaccines  
PS Claim 1; Page 1287-1288; 3271pp; English.  
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S. aureus DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against S. aureus infection. The

CC polypeptides can also be used in a kit for the immunodetection of  
CC S. aureus in a sample. S. aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S. aureus DNA sequences contained on the  
CC computer readable medium.  
SQ Sequence 2424 BP; 839 A; 302 C; 462 G; 759 T;

Query Match 1.2%; Score 36.6; DB 1; Length 2424;  
Best Local Similarity 53.1%; Pred. No. 3;  
Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 2795 AATAATCATCATCCTTTTGTGATTTAATAATTTCTAATGATTTAGACTTC 2854  
DB 1811 AATTAAATGAAGATTTCTAATGATTTAGACAAATTTGATATGCTGTTATTTAT 1870  
QY 2855 TGTAGGGGCGATATATCTAATGATATAGTACATTTAATGATTTCTGTAGGG 2914  
DB 1871 GGGTGCAGGTGATATCAAAATTCACAAATGCATATTTAGATTAATGCGATGAAAA 1930  
QY 2915 GCGATATATCTAATGATTTTACT 2941  
DB 1931 TCGGTTTATATGTTTATATATAGAT 1957

## RESULT 14

ID Q42541/c  
ID Q42541 standard; DNA; 8700 BP.

AC 15-SEP-1993 (first entry)  
DE BgIII/HpaII fragment of pTn32.  
KW epII; epIIA; epIB; epIC; epID; epIO; epIP; epidermin; derivatives; ss.  
OS Staphylococcus epidermis.  
FH Key Location/Qualifiers  
FT cds complement (1227..781)

FT /\*tag= a  
FT /note= "epII CDS"  
FT complement (6983..6366)  
FT /\*tag= b  
FT /note= "epIQ CDS"  
FT complement (8379..6994)  
FT /\*tag= c  
FT /note= "epIP CDS"  
FT 1381..4564  
FT /\*tag= d  
FT /note= "epIIA CDS"  
FT 4443..5808  
FT /\*tag= e  
FT /note= "epIC CDS"  
FT 5824..6369  
FT /\*tag= f  
FT /note= "epID CDS"  
FT 1594..4565  
FT /\*tag= g  
FT /note= "epIB CDS"

PD EP-543195-A.  
PD 26-MAY-1993.  
PF 30-OCT-1992; 118598.  
PR 31-OCT-1991; US-784234.  
PA (THOM) THOMAS GMBH KARL.  
PI Augustin J, Engelke G, Entian K, Gotz F, Jung G, Kajaleta C, Klein C,  
PI Kellner R, Kupke T, Rosenstein R, Schmitt N, Wieland B;  
PI WPI; 93-168917/21.  
DR P-PDB; R37345, R39343, R39344, R39345, R39352, R39353, R39354.  
PT Novel DNA molecule - encoding Epi B, C, D, P or Q enzymes  
PS Claim 1; Fig 9; 52pp; English.  
CC This is the nucleotide sequence of the BgIII/HpaII fragment of the  
CC Staphylococcus epidermis plasmid pTn32 which contains open reading

CC frames for epia, epib, epic, epid, epip, epiq and epiy. The proteins  
CC encoded by these are valuable reagents potentially useful in the  
CC prepn. of novel proteins such as dehydroalanine, dehydrobutyline,  
CC meso-lanthionine, 3-methyl-lanthionine and S-(2-aminovalinyl)-D-cysteine.  
SQ Sequence 8700 BP; 3280 A; 1101 C; 1066 G; 3253 T;

Query Match 1.2%; Score 36.4; DB 1; Length 8700;  
Best Local Similarity 52.7%; Pred. No. 6.6;  
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 2762 TAGCCTAGGCGCGATTTTTCACATAAATAAATATCATCTTTTGTGAT 2821  
DB 5529 TAGTCGCGACGAAGGTTGACATTAATCAATCATTTCTTATCTTATTTAAT 5470  
QY 2822 TATAAATTTTCTAATATTTAGACTTCCTGAGGGGCGATATCAATATGATA 2881  
DB 5469 TAAATATTTTATGACATTTTAAAGCCTCATTTATTAAGCCTTACGACAAAG 5410  
QY 2882 TAGTACATTATATCAATATGATTCCTGTA 2911  
DB 5409 TAATGATTCATATCCCTGATCACCGTA 5380

## RESULT 15

ID V33135 standard; DNA; 5849 BP.  
AC V33135.  
DT 07-DEC-1998 (first entry)  
DE Plasmodium berghei plasmid PSL-PL470 gene.  
KW Malaria, infection; therapy; diagnosis; vaccine; plasmid;  
KW PSL-PL470 gene; ds.  
OS Plasmodium berghei ANKA strain.  
PN WO9835057-A1.  
PD 13-AUG-1998.  
PR 05-FEB-1998; IB0212.  
PR 26-SEP-1997; AU-009481.  
PR 06-FEB-1997; AU-004953.  
PR 21-APR-1997; AU-006329.  
PA (MOLE-) INST MOLECULAR & CELL BIOLOGY.  
PA (UYSI-) UNIV SINGAPORE NAT.  
PI Kara AKU, Nelson JS, Tan TWC, Tham JM, Ting RCY;  
DR WPT: 98-447251/38.  
PT Detecting Plasmodium infection from hybridisation with  
PT extrachromosomal element - providing genus or species specific  
PS Claim 15: Page 54-59; 120pp; English.  
CC This is the nucleotide sequence of one strand of the PSL-PL470  
CC gene of the 30.7 kb extrachromosomal plasmid of plasmodium berghei.  
CC This plasmid encodes organelle-like rRNAs, tRNAs, ribosomal  
CC proteins and RNA polymerase subunits, amongst others. Plasmodium is  
CC detected in a human or animal sample by treating it, or derived  
CC nucleic acid, with a Plasmodium extrachromosomal genetic element or  
CC derived nucleic acid (A) and detecting any hybridisation. (A) can  
CC include the PSL-PL470, PLH-PBH, PRB or PMQ gene, the mitochondrial  
CC coxI gene, and nucleic acids derived from them. Also new are  
CC (A)-specific probes and primers (see V33139-56). The method is  
CC used to diagnose Plasmodium infection. Also (not claimed) the  
CC polypeptides encoded by (A) are useful as targets for drug  
CC development and for development of anti-malaria vaccines. The high  
CC degree of similarity between (A) from different species allows  
CC development of genus- or species-specific assays that result in  
CC fewer false negatives than known methods (typically 1% against 3%).  
SQ Sequence 5849 BP; 2296 A; 673 C; 557 G; 2323 T;

Query Match 1.2%; Score 36.4; DB 1; Length 5849;  
Best Local Similarity 50.0%; Pred. No. 5.4;  
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 2787 ATAAATAAATAATCATCATCTTTTGTGATATAAATTTCTAAATGTATTTT 2846  
DB 3399 ATAAATAAATTTTAAATATTTTATTTTATTAATATAAATTTTCGTAATATTTTA 3458

QY 2847 AGACTTCCTGAGGGGCGATATCAATGATATATGTCATTTATACATAATGTTTC 2906  
DB 3459 ACGTCTTCTTCATTAATAAATAAATATAGACATCCCTTTAAATTTATATATATTTAA 3518  
QY 2907 CTGTAGGGGCGATATCAATGATATTTTATGACTTCCTGAGGGGCGATATAAATAA 2966  
DB 3519 TTATATATTTTACTATATTAATTTATTAATTTATTTTAAATTAACGAAACGGAAT 3578  
QY 2967 TG 2968  
DB 3579 TG 3580

Search completed: August 6, 1999, 04:18:30  
Job time: 2462 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:58 ; Search time 3075.15 Seconds  
(without alignments)  
1913.425 Million cell updates/sec

Title: US-09-049-696-20

Perfect score: 2983  
Sequence: 1 GAAATCAGGAGAGATGTAC.....AAATGCTAAACACTGGGTA 2983

Scoring table:

Searched: 2546578 seqs, 986266752 residues

Database: EST.\*

1: em\_est1:.\*  
2: em\_est2:.\*  
3: em\_est3:.\*  
4: em\_est4:.\*  
5: em\_est5:.\*  
6: em\_est6:.\*  
7: em\_est7:.\*  
8: em\_est8:.\*  
9: em\_est9:.\*  
10: em\_est10:.\*  
11: em\_est11:.\*  
12: em\_est12:.\*  
13: em\_est13:.\*  
14: em\_est14:.\*  
15: em\_est15:.\*  
16: em\_est16:.\*  
17: em\_est17:.\*  
18: em\_est18:.\*  
19: em\_est19:.\*  
20: em\_est20:.\*  
21: em\_est21:.\*  
22: em\_est22:.\*  
23: em\_est23:.\*  
24: em\_est24:.\*  
25: em\_est25:.\*  
26: em\_est26:.\*  
27: em\_est27:.\*  
28: em\_est28:.\*  
29: em\_est29:.\*  
30: em\_est30:.\*  
31: em\_est31:.\*  
32: em\_est32:.\*  
33: em\_est33:.\*  
34: em\_est34:.\*  
35: em\_est35:.\*  
36: em\_est36:.\*  
37: em\_est37:.\*  
38: em\_est38:.\*  
39: em\_est39:.\*  
40: em\_est40:.\*  
41: em\_est41:.\*  
42: em\_est42:.\*  
43: em\_est43:.\*  
44: em\_est44:.\*  
45: em\_est45:.\*  
46: em\_est46:.\*  
47: em\_est47:.\*  
48: em\_est48:.\*  
49: em\_est49:.\*  
50: em\_est50:.\*  
51: em\_est51:.\*  
52: em\_est52:.\*  
53: em\_est53:.\*

54: em\_est22:.\*  
55: em\_est23:.\*  
56: em\_est24:.\*  
57: em\_est25:.\*  
58: em\_est26:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	652.4	21.9	716	49	A1660234	A1660234 we68902.x
C 2	648.4	21.7	734	49	A1660957	A1660957 wf20d08.x
C 3	480	16.1	501	51	A1721275	A1721275 as82h08.x
C 4	473.6	15.9	480	35	AA581198	AA581198 n03c07.x
C 5	450	15.1	490	34	AA508854	AA508854 n122h10.s
C 6	368	12.3	392	50	A1687981	A1687981 wa77f06.x
C 7	352.4	11.8	378	31	AA296955	AA296955 EST112726
C 8	336.8	11.3	646	35	AA691335	AA691335 vs14h06.x
C 9	326	10.9	479	35	AA541829	AA541829 vj01f01.x
C 10	322.2	10.8	335	51	A1721121	A1721121 as73d08.x
C 11	309	10.4	428	30	284017	284017 SS284017.P0
C 12	303.2	10.2	455	39	AA871197	AA871197 vq32c11.x
C 13	298.2	10.0	450	37	AA734088	AA734088 vs19c08.x
C 14	293.4	9.8	310	31	AA297150	AA297150 EST112734
C 15	289.8	9.7	629	34	AA518112	AA518112 v123c06.x
C 16	273.8	9.2	537	37	AA688953	AA688953 vs04h07.x
C 17	263.4	8.8	502	36	AA623217	AA623217 v114d07.x
C 18	261.8	8.8	554	36	AA623196	AA623196 v114b07.x
C 19	259.4	8.7	469	36	AA611564	AA611564 v085d05.x
C 20	248	8.3	482	37	AA711228	AA711228 vt70c11.x
C 21	241.2	8.1	255	20	D25727	D25727 HUMGSO4094
C 22	239.2	8.0	488	37	AA733415	AA733415 vt75b06.x
C 23	229.6	7.7	449	38	AA759816	AA759816 vt81b01.x
C 24	200	6.7	452	39	AA822200	AA822200 vp36d05.x
C 25	198.4	6.7	638	45	A1390317	A1390317 mx03b07.y
C 26	192.6	6.5	441	26	W41083	W41083 mc41c10.f1
C 27	190.6	6.4	301	37	AA689856	AA689856 vt62c11.x
C 28	189.8	6.4	431	30	AA238284	AA238284 my34g09.x
C 29	188.8	6.3	524	39	AA871630	AA871630 vq39a03.x
C 30	183.8	6.2	463	39	AA839323	AA839323 v040f07.x
C 31	180.4	6.0	418	35	AA596289	AA596289 v026a08.x
C 32	171.6	5.8	370	37	AA711143	AA711143 vt56d09.x
C 33	170.4	5.7	421	37	AA691586	AA691586 vs12g06.x
C 34	170.4	5.7	388	37	AA734161	AA734161 vs19g06.x
C 35	163.2	5.5	327	23	F15082	F15082 SSC6A10.P0T
C 36	162	5.4	297	49	A1641931	A1641931 vq50e03.y
C 37	159	5.3	420	49	A1641868	A1641868 vq39a03.y
C 38	158	5.3	470	43	A1197242	A1197242 u681f01.x
C 39	155.8	5.2	161	48	A1583072	A1583072 ar96a11.x
C 40	155.2	5.2	386	37	AA692521	AA692521 vt59g08.x
C 41	151.8	5.1	271	30	AA245498	AA245498 my51e11.x
C 42	149.4	5.0	427	34	AA498535	AA498535 v140f12.x
C 43	144.4	4.8	297	40	AA929159	AA929159 vq58g07.x
C 44	133.8	4.5	468	49	A1640041	A1640041 m667c10.y
C 45	132.8	4.5	372	38	AA755896	AA755896 v104f06.1

ALIGNMENTS

RESULT 1  
LOCUS A1660234 716 bp mRNA EST 10-MAY-1999  
DEFINITION we68902.x1 Soares\_Dieckgraefe\_colon\_NHCD Homo sapiens cDNA clone  
IMAGE:2346290 3' similar to TR:088826 O88826 GDB-5 PROTEIN. ;, mRNA  
ACCESSION A1660234  
NTD g4763804





OY	2629	CAGACCTCCGCCAGAGACACTAGTCTCGATGGAACGTCGTCTTGCTCATATTTCAT	2688
Db	201	CAGACTCCGCCAGAGACACTAGTCTCGATGGAACGTCGTCTTGCTCATATTTCAT	142
OY	2689	ATCAACAGCACCATTCTGGCATTTACATTTTTAAAAATTATGTGGAAGTGATAGAGA	2748
Db	141	ATCAACAGCACCATTCTGGCATTTACATTTTTAAAAATTATGTGGAAGTGATAGAGA	82
OY	2749	CTGCAGCTCTCAATFAGCTTAGGGCTAAATTTTGTGCGATTAATAAATAATCATTCAT	2808
Db	81	CTGCAGCTCTCAATFAGCTTAGGGCTCACTTTTGTCTCTATTATAACAANAATCATTCAT	22
OY	2809	CCTTTTTTTTGATTTAT	2824
Db	21	CCTTTTTTTTTATTAT	6
RESULT	4		
LOCUS	AA581198		
DEFINITION	AA581198	480 bp	mRNA
	nd38c07.r1 NCI-CCAP_Co1 Homo sapiens cDNA clone IMAGE:802572 5'		EST
	similar to TR:G1184066 G1184066 CALCIUM-ACTIVATED CHLORIDE CHANNEL		
	; , mRNA sequence.		
ACCESSION	AA581198		
NID	G2358970		
VERSION	AA581198.1	GI:2358970	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 480)		
AUTHORS	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
JOURNAL	Tumor Gene Index		
COMMENT	Unpublished (1997)		
	On May 9, 1995 this sequence version replaced gi:803001.		

```

FEATURES
source
    Contact: Robert Strausberg, Ph.D.
    Tel: (301) 496-1550
    Email: Robert.Strausberg@nih.gov
    Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael
    Emmert-Buck, M.D., Ph.D.
    cDNA Library Preparation: Life Technologies Inc., David Krizman,
    Ph.D.
    cDNA Library Arraying: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone distribution: NCI-CGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 405.
    Location/Qualifiers
        1..480
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:802572"
            /clone_1lb="NCI-CGAP_Col1"
            /tissue_type="bulk tumor"
            /lab_host="DH10B"
            /note="Organ: colon; Vector: PCMV-SPORT2; site_1: SalI;
            site_2: NotI; Cloned unidirectionally. Primer: Oligo dT."
BASE COUNT
168 a 101 c 97 g 114 t
ORIGIN

```

Query Match	15.9%	Score 473.6;	DB 35;	Length 480;
Best Local Similarity	99.2%	Pred. No. 3e-112;		
Matches 476, Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

Db	1	AAATTCACCTTATTCACATGGAGAAATACAAAGCAGTAAGATGTTCCAGCAGTATTACTGGT	60
QY	604	ACAATGTAGTAAAGAAAGTGTCAAGGAGGACAGCTGTACACCAAAAGATGCACATTCAAT	663
Db	61	ACAATGTAGTAAAGAAAGTGTCAAGGAGGACAGCTGTACACCAAAAGATGCACATTCAAT	120
QY	664	AAATTAACAGAGACTCATGTGAAAAAGGATGAGTTGTTCTTCCAAATCCCGCACAACGGAG	723
Db	121	AAATTTACAGAGACTCATGTGAAAAAGGATGAGTTGTTCTTCCAAATCCCGCACAACGGAG	180
QY	724	AAGGCTTCTATTAATGTTTGACACACATGTTGATCTATAGTTGAATTCGTACAGAACAA	783
Db	181	AAGGCTTCTATTAATGTTTGACACACATGTTGATCTATAGTTGAATTCGTACAGAACAA	240
QY	784	AACCAACAACAAAGAGCTCCAAACAAAGCAAAATCAAAAATGCAATCTTCGAGCACAATGG	843
Db	241	AACCAACAACAAAGAGCTCCAAACAAAGCAAAATCAAAAATGCAATCTTCGAGCACAATGG	300
QY	844	GAAATGATGCGGTATCTGTAGAGACTTTAAGAAACCACTCTATAGACACACACCCACCA	903
Db	301	GAAATGATGCGGTATCTGTAGAGACTTTAAGAAACCACTCTATAGACACACACCCACCA	360
QY	904	AATCCACACTCTTCATTCGTGTGCAGATTGGACAAAGAAATGTGTGTTTAGTCCCTGCACAA	963
Db	361	AATCCACACTCTTCATTCGTGTGCAGATTGGACAAAGAAATGTGTGTTTAGTCCCTGCACAA	420
QY	964	TCGTGAAGCATGGCGCACTGCTGTAACCGCCTCAATGCATGTAATCAAGCAGGCGCACTTTTC	1023
Db	421	TCGTGAAGCATGGCGCACTGCTGTAACCGCCTCAATGCATGTAATCAAGCAGGCGCACTTTTC	480
RESULT	5		
AA508854/c			
LOCUS	AA508854	490 bp	mRNA
DEFINITION	n122110.s1 NCI_CGAP_C04 Homo sapiens cDNA clone IMAGE:18-AUG-1997		
ACCESSION	AA508854		
NID	92246357		
VERSION	AA508854.1	GI:2246357	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
COMMENT	Unpublished (1997)		
	On Sep 12, 1996 this sequence version replaced gi:1397407.		

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,  
 M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
  
 Insert Length: 856 Std Error: 0.00  
 Seq primer: -40ml3 fwd. Ex from Amersham  
 High quality sequence stop: 404.  
 Location/Qualifiers  
     1. 490  
         /organism="Homo sapiens"  
         /db\_xref="taxon:9606"  
         /clone="IMAGE:968803"  
         /clone\_1lp="NCI\_CGAP\_CO4"











Db 363 AAAATGATGGGCGCTATACGACAGAGATGGGNACTWTGCGAAGAAAAGGTGGAAAAATTT 422

Or 422 AACTCA 427  
1 111  
Db 423 ATTCA 428

RESULT 12  
AA871197

LOCUS  
DEFINITION  
IMAGE:1095956 5' similar to SW:PCIC\_BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.

ACCESSION  
AA871197  
92966642  
AA871197.1 GI:2966642

VERSION  
KEYWORDS  
EST.

SOURCE  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ORGANISM

REFERENCE  
1 (bases 1 to 455)

AUTHORS  
Marra,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuqu  
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,R., Moore,B.,  
Teisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE  
JOURNAL  
COMMENT  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
On Sep 17, 1996 this sequence version replaced gi:1395394.

Contact: Marra M/Mouse EST Project  
 Mashu-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@img.lnl.gov) for further information.  
 MGI:602188  
 Seq primer: -28ml3 rev2 ET from Amerisham  
 High quality sequence stop: 414.  
 Location/Qualifiers  
 1..455

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1095956"
/clone_lib="Barstead bowel MPLB9"
/tissue_type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo (dT) primer [5',
TGTATACCAATCTGAAATGGAGCGCGCCCTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGAGATCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pRT73 vector.
Source: irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob
Barstead."

```

```

Query Match 10.28; Score 303.2; DB 39; Length 455;
Best Local Similarity 79.48; Pred. No. 3.9e+68;
Matches 359; Conservative 0; Mismatches 93; Indels 0; Gaps 0
07 1416 GAACATGGCCCTCATGTGATCTTTTGGGGCCCTTTCATCAGGAATGAGCTGTCTCA 1475
||||||| || ||||||||| || || || ||||||||| || || |||||

```

Db	4	GAACAATGGCTTTGGTGAATGCTTTGGCAGCACTGCTCCTCAGGAAATGGGGGATGCTCA	63
QY	1476	GGCGTCCATCCAGCTGTGAGATTAAGGATTAAACCTCCAGAAACCCAGTGGATGATGG	1535
Db	64	GCATCCCATCCAGCGTGGAGACAGAGGGGAGTTAAATCTCCAGATTAACCAATGGATGAATGG	123
QY	1556	CACAGTATCGTGGCAGACACCCTGGGGAAGGACACTTTGTTTCTTATCACCTGGACAAC	1595
Db	124	CTCAGTGAATCGTGACACACTGGTGGGCAAGACACCTGTTTCTTATCACCTGGACAAC	183
QY	1556	GCAGCTCCCAAAATCCTCTCTGGGATCCAGTGGAGACAGAGAAGCAAGTGGCTTGTAGT	1655
Db	184	GCATCTCTCTCAATATTATATCTGGGATCCAGCGGAGATGGAACAAATGGTTTATACT	243
QY	1656	GGACAAAACACCAAAATGGCGCTTACCTCCAAATCCCAAGCAATGCTAGGTTGGCACTG	1715
Db	244	AGACCAACCACTAAGGGGCGCTACCTCCAAAGTCCCAAGCAGCGCTAAGGTTGGCTTGTG	303
QY	1716	GAATAGAGTGTGCAAGCAAGCTCACAAACCTTGACCTGACTGACAGTCCCGGCGCTC	1775
Db	304	GAATAACAGCATTTCAAGGAGACTCACAAACCTCAACCTTGACTGACACTCCCGTGACGC	363
QY	1776	CAATGCTACCGCTGCCTCCAAATTACAGTAGCTTCCAAAAGCAACAGACACAGCAATT	1835
Db	364	AAGTGTACACTGCCTCTATTACAGTGAACCCCGGTAGTGAATAGAACAAGGGAATT	423
QY	1836	CCCCAGCCCTCTGGTAGTTTATGCAAAATATTC	1867
Db	424	CCCCAGCCCTGTACAGTGTATGCAAGCATTC	455

RESULT 13  
AA734088  
LOCUS  
DEFINITION  
AA734088 450 bp mRNA EST 07-JAN-1998  
vs19c08.1 Barstead mouse irradiated colon MPlRB7 Mus musculus CDNA  
clone IMAGE:1138670 5' similar to SW:EDIC\_BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.  
ACCESSION  
NID  
AA734088  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 450)  
Marrar,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Wetstling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMMI Mouse EST Project  
Unpublished (1996)  
On May 8, 1995 this sequence version replaced gi:801722.

Contact: Maria M/Mouse EST Project  
 WashU-HMNI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:619942  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence strop: 444.  
 Location/Qualifiers  
     1..450  
         /organism="Mus musculus"  
         /stain="FVB/N"  
         /db\_xref="taxon:10090"



Page 11

Contact: Maira M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMMGE Consortium ([info@lmgc.llnl.gov](mailto:info@lmgc.llnl.gov)) for further information.  
MG1:5528282  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 351.

1. .629

Query Match	9.78;	Score 289.8;	DB 34;	Length 629;
Post Local stmt1+4"	71.58;	Prod No. 1 30-64;		

Oy 2190 ACACAGCAGCTGTGTTTCAGCACGAACATCTCGGGAGGCCTATTGTGGCTTCT-GATG 2248  
| | | | | | | | | | | | | | | | | | | | |  
Db 14 AGACAAGCAGCTGTGCTTCAGCACGAGCATCTTCAGGGGGGATGTTGTGGCCACCAGATG 73

QY 2249 TCCCAATGCTCCCATACCTGAT-----CTCTTCCACCTGGCCAATCAGGACC 2299

Db 74 TCCCGCATGCTCCCTATTCCGCGACCATCTTCCACCCCTGTCAATCACTGACC 133

QY 2300 TGAAGCGGAATTCACGGGGCAGTCTCATTAATCTGACTTGACACGCTCCTGGGATG 2359

Db 134 TGAAGGCCAGCATCCAAGGCGAGAACCTGGTGAATCTGACGTGGACGGCTCCTGGGATG 193

```
Search completed: August 6, 1999, 13:34:12
Job time: 13050 sec
```

*This Page Blank (uspto)*

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:50 ; Search time 3075.15 Seconds

(without alignments)  
1079.549 Million cell updates/sec

Title: US-09-049-696-19

Perfect score: 1683

Sequence: 1 AACCAAGTGGTGCATCATC.....AAATGCTAACCACTGGGTA 1683

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:\*

1: em\_est1:\*

2: em\_est2:\*

3: em\_est3:\*

4: em\_est4:\*

5: em\_est5:\*

6: em\_est6:\*

7: em\_est7:\*

8: em\_est8:\*

9: em\_est9:\*

10: em\_est10:\*

11: em\_est11:\*

12: em\_est12:\*

13: em\_est13:\*

14: em\_est14:\*

15: em\_est15:\*

16: em\_est16:\*

17: em\_est17:\*

18: em\_est18:\*

19: em\_est19:\*

20: em\_est20:\*

21: em\_est21:\*

22: em\_est22:\*

23: em\_est23:\*

24: em\_est24:\*

25: em\_est25:\*

26: em\_est26:\*

27: em\_est27:\*

28: em\_est28:\*

29: em\_est29:\*

30: em\_est30:\*

31: em\_est31:\*

32: em\_est32:\*

33: em\_est33:\*

34: em\_est34:\*

35: em\_est35:\*

36: em\_est36:\*

37: em\_est37:\*

38: em\_est38:\*

39: em\_est39:\*

40: em\_est40:\*

41: em\_est41:\*

42: em\_est42:\*

43: em\_est43:\*

44: em\_est44:\*

45: em\_est45:\*

46: em\_est46:\*

47: em\_est47:\*

48: em\_est48:\*

49: em\_est49:\*

50: em\_est50:\*

51: em\_est51:\*

52: em\_est52:\*

53: em\_est53:\*

54: em\_est23:\*

55: em\_est23:\*

56: em\_est24:\*

57: em\_est25:\*

58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	652.4	38.8	716	49	A1660234 we68902.x
C 2	648.4	38.5	734	49	A1660957 wf40d08.x
C 3	480	28.5	501	51	A1721275 as82h08.x
C 4	450	26.7	490	34	AA508854 n122h10.s
C 5	368	21.9	392	50	A1687981 wa77f06.x
C 6	352.4	20.9	378	31	AA296955 EST112726
C 7	336.8	20.0	646	37	AA691335 vs14h06.x
C 8	326	19.4	479	35	AA541829 vj01f01.x
C 9	323.2	19.1	335	51	A1721121 as73d08.x
C 10	303.2	18.0	455	39	AA871197 vq32c11.x
C 11	298.2	17.7	450	37	AA734088 vs19c08.x
C 12	293.4	17.4	310	31	AA297150 EST112734
C 13	289.8	17.2	629	34	AA518112 v123c06.x
C 14	273.8	16.3	552	37	AA688953 vs04h07.x
C 15	263.4	15.7	507	36	AA623217 v114d07.x
C 16	261.8	15.6	554	36	AA623196 v114d07.x
C 17	259.4	15.4	469	36	AA611564 v085d05.x
C 18	248	14.7	482	37	AA711228 vt70c11.x
C 19	241.2	14.3	255	20	D25727 HUM604094
C 20	239.2	14.2	488	37	AA733415 vt75b06.x
C 21	229.6	13.6	449	38	AA759816 vt81b01.x
C 22	200	11.9	452	39	AA822200 vp36d05.x
C 23	192.6	11.4	441	26	W41083 mc41c10.f1
C 24	190.6	11.3	301	37	AA689856 vt62c11.x
C 25	189.8	11.3	431	30	AA238284 my34g09.x
C 26	188.8	11.2	524	39	AA871630 vq39a03.x
C 27	183.8	10.9	463	39	AA839323 v040f07.x
C 28	180.4	10.7	418	35	AA596289 v026a08.x
C 29	171.6	10.2	370	37	AA711143 vt56d09.x
C 30	170.4	10.1	421	37	AA691586 vs12g06.x
C 31	170.4	10.1	388	37	AA734161 vs19g06.x
C 32	162	9.6	297	49	A1641931 vq50e03.y
C 33	159	9.4	420	49	A1641868 vq39a03.y
C 34	158	9.4	470	43	A1197242 u081f01.x
C 35	155.8	9.3	161	48	A1582072 at96a11.x
C 36	155.2	9.2	386	37	AA692521 vt59g08.x
C 37	151.8	9.0	271	30	AA245498 mv51e11.x
C 38	144.4	8.6	297	40	AA929159 vq58g07.x
C 39	133.8	8.0	468	49	A1640041 ms67c10.y
C 40	130.4	7.7	207	39	AA869318 vq50e03.x
C 41	129.8	7.7	205	50	A1675991 w080g07.x
C 42	124.2	7.4	466	29	AA170458 ms67c10.x
C 43	115.6	6.9	175	47	A1506262 vq50e03.x
C 44	115	6.8	243	26	W20635 mc16f08.f1
C 45	103	6.1	435	33	AA443258 aa14d09.f

#### ALIGNMENTS

RESULT 1

LOCUS A1660234 716 bp mRNA

DEFINITION we68902.x1 Soares Dieckgraebe.colon\_NHCD Homo sapiens cDNA clone

IMAGE:2346290 3' similar to TR:088826 O88826 GDB-5 PROTEIN. ; mRNA

ACCESSION A1660234

NTD g4763804

EST 10-MAY-1999

VERSION A1660234.1 GI:4763804  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 716)  
 AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2949219.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 387.  
 Location/Qualifiers  
 1..716  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2346290"  
 /tissue\_type="Soares\_Dieckgraefe\_colon\_NHUC"  
 /tissue\_type="colonic mucosa from 3 patients with Crohn's  
 disease"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a  
 modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTACCAATCTGAAGTGGAGCGCCGCTCTTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified p773 vector. Library  
 went through one round of normalization. Tissue samples  
 provided by Dr. Brian Dieckgraefe (Washington University,  
 dieck@im.wustl.edu); colonic mucosa represents a range of  
 disease involvement from moderate to severe Crohn's  
 disease; samples include both perforating (fistulas) and  
 non-perforating samples. Library constructed by Bento  
 Soares and M. Fatima Bonaldo.  
 BASE COUNT 220 a 123 c 140 g 233 t  
 ORIGIN  
 Query Match 38.8%; Score 652.4; DB 49; Length 716;  
 Best Local Similarity 97.6%; Pred. No. 3.4e-158;  
 Matches 662; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
 QY 1006 CGGAATTCAGGGGGGCGTCTCATTTATGACTGACAGCCTCGGGGATGATTATG 1065  
 DB 715 CCGGAATTCAGTGGAGCTCATTTATTCACCTGGACAGCTCCGCGATGATTATG 656  
 QY 1066 ACCATGGAAGCTCAACAAGTATTCATTGGAATAGTACAAAGTATTCATGACAG 1125  
 DB 655 ACCATGTAACGCTCAACAAGTATTCATTGGAATAGTACAAAGTATTCATGACAG 596  
 QY 1126 ACAAGTCAATGAATCTCTCAAGTATGATCTGCTCATCCCAAGAGCAACT 1185  
 DB 595 ACAAGTCAATGAATCTCTCAAGTATGATCTGCTCATCCCAAGAGCAACT 536  
 QY 1186 CTGAGGAAGCTTTTGTGTTAAACCAAAACATTACTTTGAAATGGCACAGATCTTT 1245  
 DB 535 CTGAGGAAGCTTTTGTGTTAAACCAAAACATTACTTTGAAATGGCACAGATCTTT 476  
 QY 1246 TCATTGCAATCAGGCTGTATAGGTCGATCGAATCGAATATCCAAACATTGAC 1305  
 DB 475 TCATTGCAATCAGGCTGTATAGGTCGATCGAATATCCAAACATTGAC 416  
 QY 1306 GAGTATCTTTGTTTATCTCCACAGACTCCGACAGACACTAGTCTGATGAAAGT 1365  
 |||||12|||||

DB 415 GAGTATCTTTGTTTATCTCCACAGACTCCGACAGACACTAGTCTGATGAAAGT 356  
 QY 1366 CTGCTCCCTTGTCTTAATTCATATCAACAGACACCACTCCGCGATTCACATTTTAAAA 1425  
 DB 355 CTGCTCCCTTGTCTTAATTCATATCAACAGACACCACTCCGCGATTCACATTTTAAAA 296  
 QY 1426 TTATGTGAATGTGATAGAGAACTGCAGCTGTCTAATAGCCTAGGCTGAATTTTGTCA 1485  
 DB 295 TTATGTGAATGTGATAGAGAACTGCAGCTGTCTAATAGCCTAGGCTGAATTTTGTCA 236  
 QY 1486 GATAAATTAATAATCATTCATCTTTTGTGATATATAAATTTCTAATAATGATTT 1545  
 DB 235 GATAAATTAATAATCATTCATCTTTTGTGATATATAAATTTCTAATAATGATTT 176  
 QY 1546 TAGACTCCCTGTAGGGGCGATATACATAATGATATAGTACATTTTACTAATGATTT 1605  
 DB 175 TAGACTCCCTGTAGGGGCGATATACATAATGATATAGTACATTTTACTAATGATTT 116  
 QY 1606 CCTTAGGGGGCGATATACATAATGATTTTACACTTCCTGTAGGGGCGATATAATAA 1665  
 DB 115 CCTTAGGGGGCGATATACATAATGATTTTACACTTCCTGTAGGGGCGATATAATAA 56  
 QY 1666 ATGCTAAACAACCTGGTA 1683  
 DB 55 ATGCTAAACAACCTGGTA 38  
 RESULT 2  
 A1660957/c LOCUS  
 DEFINITION w120608.x1 Soares Dieckgraefe\_colon\_NHUC Homo sapiens cDNA clone  
 IMAGE:2351151 3' similar to TR:088826 O88826 GDB-5 PROTEIN.; mRNA  
 sequence.  
 ACCESSION A1660957  
 NID 94764540  
 VERSION A1660957.1 GI:4764540  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 734)  
 AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Mar 20, 1998 this sequence version replaced gi:2980033.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 477.  
 Location/Qualifiers  
 1..734  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="X"  
 /clone="IMAGE:2351151"  
 /tissue\_type="Soares\_Dieckgraefe\_colon\_NHUC"  
 /tissue\_type="colonic mucosa from 5 ulcerative colitis  
 patients"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a  
 modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTACCAATCTGAAGTGGAGCGCCGCTCTTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified p773 vector. Library  
 went through one round of normalization. Tissue samples

provided by Dr. Brian Dieckgraefe (Washington University, dieckgr.b@wustl.edu); colonic mucosa represents a range of disease involvement from mild cryptitis to severe ulceration, fibrosis, and degeneration. Library constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 200 a 143 c 159 g 228 t 4 others  
 ORIGIN

Query Match 38.5%; Score 648.4; DB 49; Length 734;  
 Best Local Similarity 95.8%; Pred. No. 3.7e-157;  
 Matches 707; Conservative 0; Mismatches 25; Indels 6; Gaps 4;

QY 777 AGACGAGAGTGTATACCCAGAGAGTGTATACCTGGCTGGATTGAGAAT 836  
 DB 734 AGACGAGAGTGTATACCCAGAGAGTGTATACCTGGCTGGATTGAGAAT 675

QY 837 GATGATATCAATGATCCAGACAGACCTGAATTAAGATGATGTCAACACAG 896  
 DB 674 GATG--ATATCAATGATCAACAGACCTGAATTAAGATGATGTCAACACAG 619

QY 897 CAAGTGTGTTTCAGACAAATCTCGGAGGCTCATTTGTGGCTTCTGATGCCAAT 956  
 DB 618 CAAGTGTGTT--CAGACAAATCTCGGAGGCTCATTTGTGGCTTCTGATGCCAAT 561

QY 957 GCTCCATACCTGATCTCTTCCACCTGGCCAAATACAGCACTGAAGGGGAATTCAC 1016  
 DB 560 GCTCCATACCTGATCTCTTCCACCTGGCCAAATACAGCACTGAAGGGGAATTCAC 501

QY 1017 GGGGAGAGTCTCTTATCTGACTTGGACAGCTCTGGGGATATATGACATGGAACA 1076  
 DB 500 GGGGAGAGTCTCTTATCTGACTTGGACAGCTCTGGGGATATATGACATGGAACA 441

QY 1077 GCTCACAAGATATATCATTCGATTAAGTATGATCTGATCTCAGAGCAAGTTCAAT 1136  
 DB 440 GCTCACAAGATATATCATTCGATTAAGTATGATCTGATCTCAGAGCAAGTTCAAT 381

QY 1137 GAATCTCTTCAAGTAACTACTGCTCTCATGCCAAGAGCAACTGTGAGGAATC 1196  
 DB 380 GAATCTCTTCAAGTAACTACTGCTCTCATGCCAAGAGCAACTGTGAGGAATC 321

QY 1197 TTTTGTAAACCAAAAACATTTGTAAGTAAAGGACAGATCTTTGATGCTAT 1256  
 DB 320 TTTTGTAAACCAAAAACATTTGTAAGTAAAGGACAGATCTTTGATGCTAT 261

QY 1257 CAGGCTGTGATGAAGTGCATGTAATCAGAAATATCCAACTGACAGATCTTTG 1316  
 DB 260 CAGGCTGTGATGAAGTGCATGTAATCAGAAATATCCAACTGACAGATCTTTG 201

QY 1317 TTTATCTCTCAGACACTCGCCAGAGACACTAGTCTGATGAAAAGTCTCTTGT 1376  
 DB 200 TTTATCTCTCAGACACTCGCCAGAGACACTAGTCTGATGAAAAGTCTCTTGT 141

QY 1377 CCTAATATCATATCAACAGCACTCTCGGATTCACATTTTAAATAATAGTGAAG 1436  
 DB 140 CCTAATATCATATCAACAGCACTCTCGGATTCACATTTTAAATAATAGTGAAG 81

QY 1437 TGGATGAGGAAGTGCAGTGTCAATAGCTAGAGGCTGAATTTTGTGAGATTAATAA 1496  
 DB 80 TGGATGAGGAAGTGCAGTGTCAATAGCTAGAGGCTGAATTTTGTGAGATTAATAA 21

QY 1497 TAAATCATTCATCTTTT 1514  
 DB 20 TAAATCATTCATCTTTT 3

RESULT 3  
 A1721275 501 bp mRNA EST 10-JUN-1999  
 LOCUS as82h08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
 DEFINITION IMAGE:2335263.3 similar to SW:ECIC\_BOVIN P54281 EPTHELIAL  
 CHROMIDE CHANNEL PROTEIN ; mRNA sequence.  
 ACCESSION A1721275

NID 95038531  
 VERSION A1721275.1 GI:5038531  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 501)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, J., Getzel, G., Jost, S., Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R. WashU-NCI human EST Project  
 Unpublished (1997)  
 On Jun 22, 1998 this sequence version replaced gi:3246918.

TITLE JOURNAL  
 COMMENT

FEATURES  
 source  
 1..501  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="xp11.2"  
 /clone="IMAGE:2335263"  
 /clone\_1ib="Barstead colon HPLRB7"  
 /sex="male"  
 /dev\_stage="adult, age 25"  
 /lab\_host="DH10B (phage resistant)"  
 /note="Organ: colon; Vector: pRT30-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA was primed with a Not I - Oligo(dT) primer (5' GTTACGATCTGAGTGGAGGAGCCGCCCTTTTATTTTATTTTATTTT 3'); double-stranded cDNA was ligated to Eco RI adaptors (5' AATCAGTAAAT 3' and 5' ATTACATG 3'), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT30 vector. Library constructed by Bob Barstead."

BASE COUNT 158 a 83 c 106 g 154 t  
 ORIGIN

Query Match 28.5%; Score 480; DB 51; Length 501;  
 Best Local Similarity 98.0%; Pred. No. 9.8e-114;  
 Matches 486; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1029 ATTAATCTGACTGGACACTCTGGGATGATATGACCATGGAACACTCACAAGTAT 1088  
 DB 501 ATTAATCTGACTGGACACTCTGGGATGATATGACCATGGAACACTCACAAGTAT 442

QY 1089 ATCATCGAATAGTCAAGTATTTCTGATCTCAGAGCAAGTCAATGATCTTCAA 1148  
 DB 441 ATCATCGAATAGTCAAGTATTTCTGATCTCAGAGCAAGTCAATGATCTTCAA 382

QY 1149 GTGAATACTAGTCTCTCATCCAAAGAGCAAGCACTGAGGAAGTCTTTGTATAA 1208  
 DB 381 GTGAATACTAGTCTCTCATCCAAAGAGCAAGCACTGAGGAAGTCTTTGTATAA 322

QY 1209 CCAGAAATATCTTTTAAATGACAGATCTTTATGCTATAGAGCTGTGAT 1268  
 DB 321 CCAGAAATATCTTTTAAATGACAGATCTTTATGCTATAGAGCTGTGAT 262

QY 1269 AAGGTGATCTGAAATCAGAAATATCAACATGACAGAGTATCTTTGTTATTCCTCA 1328  
 DB 261 AAGGTGATCTGAAATCAGAAATATCAACATGACAGAGTATCTTTGTTATTCCTCA 202

QY 1329 CAGACTCCGCGAGAGACACCTAGTCTGTGATGAAAGCTGCTCTCTTATATTCAT 1388  
 |||||||  
 DB 201 CAGACTCCGCGAGAGACACCTAGTCTGTGATGAAAGCTGCTCTCTTATATTCAT 142  
 |||||||  
 QY 1389 ATCAACAGACACCTAGTCTGTGATGAAAGCTGCTCTCTTATATTCAT 1448  
 |||||||  
 DB 141 ATCAACAGACACCTAGTCTGTGATGAAAGCTGCTCTCTTATATTCAT 82  
 |||||||  
 QY 1449 CTGACGCTGCTGATGAGGCTGATGAGGCTGATGAGGCTGATGAGGCTGATGAGGCT 1508  
 |||||||  
 DB 81 CTGACGCTGCTGATGAGGCTGATGAGGCTGATGAGGCTGATGAGGCTGATGAGGCT 22  
 |||||||  
 QY 1509 CCTTTTGTGATTTAT 1524  
 |||||||  
 DB 21 CCTTTTGTGATTTAT 6

RESULT 4  
 AA508854/c 490 bp mRNA EST 18-AUG-1997  
 LOCUS n122h10.s1 NCI\_CGAP\_C04 Homo sapiens cDNA clone IMAGE:968803 3',  
 DEFINITION mRNA sequence.  
 ACCESSION AA508854  
 NID 92246357  
 VERSION AA508854.1 GI:2246357  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 490)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1397407.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,  
 M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/dbfp/image/image.html

Insert Length: 856 Std Error: 0.00  
 Seq Primer: -40m3 fwd. ET from Amersham  
 High quality sequence stop: 404.

## FEATURES

source

1..490  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="968803"  
 /clone\_lib="NCI\_CGAP\_C04"  
 /sex="pooled"  
 /tissue\_type="colon"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. 1st strand cDNA was prepared from pooled colon  
 tumor tissue, and was then primed with a Not I - Oligo(dt)  
 primer. Double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 This library is not normalized. Library constructed by  
 Bento Soares and M. Fatima Bonaldi."

BASE COUNT  
 ORIGIN .....

155 a 82 c 89 g 164 t

Query Match 26.7%; Score 450; DB 34; Length 490;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-106;  
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1234 GCACGATCTTTTCACTTCTATTCAGGCTGTGATAGGCTGATGATGATGATGATGATGAT 1293  
 |||||||  
 DB 490 GCACGATCTTTTCACTTCTATTCAGGCTGTGATAGGCTGATGATGATGATGATGATGAT 431  
 |||||||  
 QY 1294 CCAATGTCAGAGATATCTTGTATTCCTCCACAGACTCCGCGACAGACACTAGTCT 1353  
 |||||||  
 DB 430 CCAATGTCAGAGATATCTTGTATTCCTCCACAGACTCCGCGACAGACACTAGTCT 371  
 |||||||  
 QY 1354 CTGATGAAGCTGCTGCTCTGCTCTTAATATTCATATCAGACACCACTTCTGCAATTC 1413  
 |||||||  
 DB 370 CTGATGAAGCTGCTGCTCTGCTCTTAATATTCATATCAGACACCACTTCTGCAATTC 311  
 |||||||  
 QY 1414 ACATTTTAAATTTATGAGAGAGATGATGAGAGAGATGATGAGAGAGATGATGAGAGAGAT 1473  
 |||||||  
 DB 310 ACATTTTAAATTTATGAGAGAGATGATGAGAGAGATGATGAGAGAGATGATGAGAGAGAT 251  
 |||||||  
 QY 1474 GAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1533  
 |||||||  
 DB 250 GAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191  
 |||||||  
 QY 1534 TAAATGATTTTACATCTCTGATGAGGCGGATATGATGATGATGATGATGATGATGATGAT 1593  
 |||||||  
 DB 190 TAAATGATTTTACATCTCTGATGAGGCGGATATGATGATGATGATGATGATGATGATGAT 131  
 |||||||  
 QY 1594 ACTAATGATTTCTGATGAGGCGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1653  
 |||||||  
 DB 130 ACTAATGATTTCTGATGAGGCGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 71  
 |||||||  
 QY 1654 CGATTAATTAATTAATGCTTAACACTGGGTA 1683  
 |||||||  
 DB 70 CGATTAATTAATTAATGCTTAACACTGGGTA 41  
 |||||||

RESULT 5  
 A1687981/c 392 bp mRNA EST 27-MAY-1999  
 LOCUS A1687981  
 DEFINITION wa77f06.x1 Soares.NFL.T.GBC.S1 Homo sapiens cDNA clone  
 IMAGE:2302211 3' similar to TR:088826 O88826 GDB-5 PROTEIN. ;, mRNA  
 sequence.  
 ACCESSION A1687981  
 NID 94899275  
 VERSION A1687981.1 GI:4899275  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 392)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2947973.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40UP from Gibco.

## FEATURES

source

1..392  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="2302211"  
 /clone\_lib="Soares.NFL.T.GBC.S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with



a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbH19w, testis NHT, and B-cell NCI-GAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as a tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 125 a 69 c 67 g 131 t

Query Match 21.9%; Score 368; DB 50; Length 392;

Best Local Similarity 98.5%; Pred. No. 7.5e-85; Matches 382; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1296 AACATGCGACGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACCTAGTCCCT 1355  
 DB 392 AACATGCGACGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACCTAGTCCCT 333  
 QY 1356 GATGAACGCTGCTCCTGCTCTTCTAATATCATATCAACAGACCAATTCCTGCAATTCAC 1415  
 DB 332 GATGAACGCTGCTCCTGCTCTTCTAATATCATATCAACAGACCAATTCCTGCAATTCAC 273  
 QY 1416 ATTTTAAATATATGAGAGATGAGAGAACTGCAGCTGTCATATGCTTACCTGCTGA 1475  
 DB 272 ATTTTAAATATATGAGAGATGAGAGAACTGCAGCTGTCATATGCTTACCTGCTGA 213  
 QY 1476 ATTTTGTGAGAT 1335  
 DB 212 ATTTTGTGAGAT 153  
 QY 1536 AATGTATTTTACTCTCTGAGGGGGGAGATATATATATATATATATATATATATATATAT 1595  
 DB 152 AATGTATTTTACTCTCTGAGGGGGGAGATATATATATATATATATATATATATATATAT 94  
 QY 1596 TAAATGATTTCTGAGGGGGGAGATATATATATATATATATATATATATATATATAT 1655  
 DB 93 TAAATGATTTCTGAGGGGGGAGATATATATATATATATATATATATATATATATATAT 34  
 QY 1656 ATTAAT 1683  
 DB 33 ATTAAT 6

RESULT 6  
 LOCUS AA296955 378 bp mRNA EST 18-APR-1997  
 DEFINITION EST112726 Colon 1 Homo sapiens cDNA 5' end, mRNA sequence.  
 ACCESSION AA296955  
 NID 91949515  
 VERSION AA296955.1 GI:1949515  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 378)  
 AUTHORS Adams,M.D., Keriavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodex,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uteckback,T.R., Weidman,J.F., Li,Y., Bedarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A.,

He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,T.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M., and Venter,J.C.  
 Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl.), 3-174 (1995)

JOURNAL MEDLINE 96026280  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1288227.  
 Other ESTs: THCL67738  
 Contact: Kerlavage, AR  
 Bioinformatics

The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M3 Reverse.

FEATURES  
 source location/Qualifiers

1..378  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (host):127834"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Colon 1"  
 /dev\_stage="adult"  
 /note="Organ: colon; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"  
 BASE COUNT 112 a 79 c 64 g 119 t 4 others  
 ORIGIN

Query Match 20.9%; Score 352.4; DB 31; Length 378;  
 Best Local Similarity 96.8%; Pred. No. 7.9e-81;  
 Matches 368; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 1143 CTTCAGTGAATATCTACTCTCTCATCCCAAGAGCCAACTGTGAGAGCTTTTGG 1202  
 DB 1 CTTCAGTGAATATCTACTCTCTCATCCCAAGAGCCAACTGTGAGAGCTTTTGG 60  
 QY 1203 TTTAAACGAAAACATCTTTTGAATATGACAGATCTTTTCTATCTATGAGCT 1262  
 DB 61 TTTAAACGAAAACATCTTTTGAATATGACAGATCTTTTCTATCTATGAGCT 120  
 QY 1263 GTTGATAGGTGATCTGAATATGAGAAATATCCACATTCACAGAGATCTTTGTTATT 1322  
 DB 121 GTTGATAGGTGATCTGAATATGAGAAATATCCACATTCACAGAGATCTTTGTTATT 180  
 QY 1323 CCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAAAGCTCTGCTGCTTAAT 1382  
 DB 181 CCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAAAGCTCTGCTGCTTAAT 240  
 QY 1383 ATTCATATCAACAGACCACTCTGCGATTCACATTTTAAATATATGGAAGTGAGATA 1442  
 DB 241 ATTCATATCAACAGACCACTCTGCGATTCACATTTTAAATATATGGAAGTGAGATA 300  
 QY 1443 GGAGACTGACAGTGCATATAGCTTGGGCTGATATTTTGTGAGATTAATTAATATC 1502  
 DB 301 GGAGACTGACAGTGCATATAGCTTGGGCTGATATTTTGTGAGATTAATTAATATC 358  
 QY 1503 ATTCATCTTTTTTTTGGATT 1522  
 DB 359 ATTCATCTTTTTTTTGGATT 378

RESULT 7  
 LOCUS AA691335 646 bp mRNA EST 16-DEC-1997  
 DEFINITION vs14h06.r1 Barstead mouse irradiated colon MFLRB7 Mus musculus cDNA  
 clone IMAGE:1138235 5' similar to SW-ECIC\_BOVIN P54281 EPITHELIAL



	Query Match	19.44:	Score 326:	DB 35:	Length 479;
	Best Local Similarity	80.96:	Pred. NO. 5.4e-74;		
	Matches 380:	Conservative	0:	Mismatches 90;	Indels 0:
				Gaps	
Qy	376	CTACTCTCCAATCCCAAGCATTGCTAAGTTGGCACTTGGAAATACAGTCTGCACAA	435		
Db	5	CCAACCTCCAGTCCCAAGCAGCGCTAAGGTGGCTTTGGAAATACAGCATTCACGCA	64		
Qy	436	GCTACAAACCTTGACCTGACTGTCAAGTCCGCTGGTCCAAATGCAACCTGGCTCAA	495		
Db	65	GCTCACAAGCTCTCACTTGACTGTCACTCCCTGCAAGCAAGTGTACCTGCTCTCTA	124		
Qy	496	TTACAGTACTTCCAAAAGCAAGACACACAGCAAAATTCGCCAGCCCTGTGGTATT	555		
Db	125	TTACAGTAGCCCCGGTAGTAGTAATGAAGACACAGGAAATTCGCCAGCCCTGTAAACATGT	184		
Qy	556	ATGCAAAATATTCGCACAGACGCTCCCAATTCACAGGCGCAGTGTACAGCCCTGATTG	615		
Db	185	ATGCAAGCATTCGCCAAGACGCTTGCTATTCTCAGGCGCAGGTACACGCTTGATTG	244		

RESULT	9
LOCUS	A1721121/c
DEFINITION	A1721121 335 bp mRNA EST 10-JUN-1999
ACCESSION	U03608.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
VERSION	IMAGE:2334351 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA
KEYWORDS	sequence.
SOURCE	A1721121
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;
	Eutheria: Primates: Catarrhini: Homnidae: Homo.
	1 (bases 1 to 335)
	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
	Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maiza, M.,
	Martin, J., Moore, B., Schellenberg, K., Stepec, M., Tan, F.,
	Thelsting, B., White, T., Wylie, T., Waterston, R. and Wilson, R.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewartson.wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from Gibco  
High quality sequence drop: 312.

```

FEATURES
SOURCE
location/Qualifiers
1..335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2334351"
/clone_1lp="Barstead colon HPLRb7"
/sex="male"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/notes="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACGACATCTGAAGTGGGACGGCCGCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[5', AATTCACCTAGTAT 3' and 5', ATTACTAGG 3'], digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pT73 vector. Library constructed by Bob
Barstead."
98 a 57 c 71 g 109 t

```

Query Match	19.1%;	Score 322.2;	DB 51;	Length 335;
Best Local Similarity	97.0%;	Pred. No. 4,8e-73;		
Matches 327;	Conservative	0;	Mismatches 8;	Indels 0; Gaps
Oy	1176 GAACCCAACTCTGAGGAAGCTTTTGGTTTAAACCAGAAACATTACTTTTGAAAATGC	1235		
Dd	335 GAACCCAACTGTAGGAAAGCTTTTGGTTTAAACCAGAAACATTACTTTTGAAAATGCC	276		
Oy	1236 ACAGATCTTTTCATTGCTATTCAGGCTGTTGATAGGTCGATCTGAAATCAGAAATATCC	1293		
Dd	275 ACAGATCTTTTCATTGCTATTCAGGCTGTTGATAGGTCGATCTGAAATCAGAAATATCC	216		
Oy	1296 AACCTTCAGAGAGATCTTGGTTTATTCCTCCACAGACTCGCGCAGAGACCTGATGCT	1355		
Dd	215 AACCTTCAGAGAGATCTTGGTTTATTCCTCCACAGACTCGCGCAGAGACCTGATGCT	156		
Oy	1356 GATAAAGCTGTGCTCTCTTCTCTTAATATTCATATCAACAGCACCATTCCTGGCATTCAC	1415		
Dd	155 GATAAAGCTGTGCTCTCTCTCTTAATATTCATATCAACAGCACCATTCCTGGCATTCAC	96		
Oy	1416 ATTTTAAAAATTANGTGGAAAGTGGATNAGGAACCTGCAGCTGTCAATAGGCTGAGA	1475		
Dd	95 ATTTTAAAAATTANGTGGAAAGTGGATNAGGAACCTGCAGCTGTCAATAGGCTGAGA	36		
Oy	1476 ATTTTGTGAGATTAATAATAATTAATCATTCATCC	1510		
Dd	35 ATTCCCATCAGATTAATAATAATTAATCATTCATCC	1		
RESULT 10				
AA871197				
LOCUS	AA871197	455 bp	MRNA	
DEFINITION	vq32c11.f1 Barsted bowel MPRB9 Mus musculus CDNA clone	EST	16-MAR-1998	
	IMAGE:1095956 5' similar to SW:ECIC_BOVIN P54281 EPITHELIAL			
	CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.			
ACCESSION	AA871197			

NID	q2966642
VERSION	AA871197.1
KEYWORDS	GI:2966642
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eutheria; Rodentia; Sciurognathi; Muridae; Mammalia; 1 (bases 1 to 455)
AUTHORS	Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thaising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R.
TITLE	The Mashu-HMM Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1395394.

Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LINTL ; contact the  
IMAGE Consortium (info@image.lnll.gov) for further information.  
MGI:602188  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 414.  
Location/Qualifiers  
1. .435

```

source
1. .455
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image:1095956
/clone_lib="Barstead bowel MPLB9"
/tissue_type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: pRT73D-Pac (pharmacia) with a modified
polylinker. Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACGATCTGAGGTGGAGCGGCCCTTTTCTTTTCTTTTCTTTT
3']; Double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGCATCTTGG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pRT73 vector.
Source irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob
Barstead."

```

Query Match	Similarity	18.0%	Score 303.2	DB 39	Length 455
Best Local	Similarity	79.4%	Pred. No. 4,1e-68		
Matches	359	Conservative	0	Mismatches 93	Indels 0
				Gaps	0
Qy	116	GACCAATGGCCATCATGATGATCTTTGGGGCCCTTTCATCAGGAATAGAGCTGTCTCA	175		
Db	4	GACCAATGGCTCTGTTGATGGCTTTCACACACTCTCTCAGGAAATCGCGCATCTCCA	63		
Qy	176	GAGTCATCCACAGCTTGAGAGTAAGGATTAACCTCCGAAACAGCAGTGGATGAATG	235		
Db	64	GCATTCATCCACGCTGGAGAGCAGGGGAATTATCTCCGAATTAACCAATGGATGATATG	123		
Qy	236	CACAGTATCGTGAGACAGCACCGTGGAAAGAGACTTTGTTTATACCTGGACAAC	295		
Db	124	CTCAGTAGATCGTGAGACAGCTCGGTGGGCAAGACACCTGTTCTTATACCTGGACAAC	183		
Qy	296	GGAGCTCCCCAAATCCCTCTCTGGGATCCCAAGTGAACAGAAAGCAAGGAGGCTTTAGT	355		
Db	184	GGATCTCTCTACAAATTTTATCTGGATCCCAAGGAGTGAACAAATGGTTTATCT	243		

QY	356	GGACAAAAACACAAATGGGCGCTACGTCCAAATCCGAGCATTTCTAAGGTGGCAATTTG	415
Db	244	AGACACAAACGACTAAGGTGGCGCTACGTCCAAATGCCAGGACGGCTAAGGTTGGCTTTTG	303
QY	416	GAAATACAGTGTGCACAGCAAGCTCACAAACCTTAACCTACTGTACGTCCCGTGGCTC	475
Db	304	GAAATACAGCATTTCAAGGAGACTCACAGACTCTACCTTGCTATCAGTCCCGTGGAGC	365
QY	476	CAATGCTACCGCTGGCTCCCAATTACAGTGACTTCCAAACGAAACAGGACACACAGCAATT	535
Db	364	AAGTCTCACTGCTGCTCTATTACAGTGATCCCGGTAGTGATTAAGAACACAGGAAATT	422
QY	536	CCCCAGCCCTGGTAGTTATGCAATATATTC	567
Db	424	CCCCAGCCCTGTAAACAGTGTATGCAGAGATTTC	455

RESULT	11		
LOCUS	AA734088		
DEFINITION	AA734088 450 bp mRNA EST 07-JAN-1998		
	v919c08.r1 Barstead cloned colon MRLB7 Mus musculus CDNA		
	clone IMAGE:1138670 5' similar to SW:EDIC_BOVIN P54281 EPITHELIAL		
	CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.		
ACCESSION	AA734088		
NID	92755755		
VERSION	AA734088.1	GI:2755755	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eukaryota; Vertebrata; Chordata; Craniata; Vertebrata; Mammalia;		
	Eutheria; Murina; Muridae; Mus.		

REFERENCE  
AUTHORS  
1 (bases 1 to 450)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Gessel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, B.,  
Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,  
Thaisang, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE  
The Mashu-HMI Mouse EST Project  
JOURNAL  
Unpublished (1996)  
COMMENT  
On May 8, 1995 this sequence version replaced g1:801722.

TITLE	JOURNAL	COMMENT
The WashU-HHMI Mouse EST Project Unpublished (1996)		On May 8, 1995 this sequence version replaced g1:801722.
Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:619942		Possible reversed clone: similarity on wrong strand Seq primer: -28ml3 rev2 ET from AmerSham High quality sequence stop: 444. Location/Qualifiers 1..450
FEATURES		SOURCE

```

/clone="IMAGE:1138670"
/clone_lib="Barstead mouse irradiated colon MPLRB"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker, site_1: EcoRI, site_2: NotI. Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
[5'GTGTACGATCTGAAGTGGAGCGCCGCTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (AATTCGATCTTG), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pRT3
vector. Library constructed by Bob Barstead. "

```



TITLE The WashU-HMI Mouse EST project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1288442.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LINDL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:525282  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 351.

FEATURES	Location/Qualifiers
source	1. .629

```

/strain="FVB/N"
/db_xref="taxon:10090"
/clone IMAGE:904618"
/clone_lib="Barstead mouse irradiated colon MPLRB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="vector: pRT73D-Pac (pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
15'GTTTACGATCTGAGATGGAGCGCCGCCCTTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors (AATTCGATCCCTTG), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pRT73
vector. Library constructed by Bob Barstead."

```

Db	374	ATATGTCATTGCTATCCAGGCTGNGATTAAGTCCAAATCTGAATCAAGAAATCTCCAAACA	433
QY	1300	TTGCACGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCACAGACACTGATCTGATG	1359
Db	434	TTGCACGGGTGTGTGTGTTCATCCCGCTCAG-----GAGCGGCCCATTTCCCGTAG	484
QY	1360	AAACGCTGCTCTGTGCTCTAATATTCATATCAACAGCACCATTCCTGGCATTCACATTT	1419
Db	485	ACTCAACTTCCCTGTGTCTGTGCATCAGAGATCAACAGCACCATTTCTGGCATCCAGTGC	544
QY	1420	TAAAAATTATGTGGAAGTGGATAGAGAGACAGCTGTCAATAGCCTAAGGCTCAATTT	1479
Db	545	TGAATATATATGTGGAAGATGGCTAGGGGAAATGCAAGTACACTAAGGTTTCACTGGAAAT	604
QY	1480	T 1480	
Db	605	T 605	

RESULT	14
AA688953	
LOCUS	552 bp mRNA EST 12-DEC-1997
DEFINITION	V004h07.r1 Barstaden mouse irradiated colon MPlRB7 Mus musculus cDNA clone IMAGE:113727 5' similar to SW-EGIC_BOVIN P54281 EPITHELIAL CHLORIDE CHANNEL PROTEIN ; mRNA sequence.
ACCESSION	AA688953
NID	92678382
VERSION	AA688953.1 GI:26783382
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE	1 (bases 1 to 552)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Gelsel,S., Knocabs,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelsting,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Nov 6, 1997 this sequence version replaced g1:930439.
FEATURES	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:618549 Seq primer: -28mJ rev2 ET from Amersham High quality sequence stop: 509.
source	location/Qualifiers 1..552

QY 1060 ATTATGACCATGGAAAGCTCAAGAATATTCATTGCAGTAATAGTACAAGATTTCTTGATC 111  
| | | | | | | | | | | | | | | | | | | | | |  
Db 194 ACTTAGCACCGGAGAGAGCTTCCAACTACATCATTCGCCAATGAGCACCAAGTATGCTTGATC 253  
| | | | | | | | | | | | | | | | | | | | | |  
QY 1120 TCAGAGACAGTGTCAATGAATCTCTTCAAGTAATACTACTGCTCTCATCCCAAAGAG 117  
| | | | | | | | | | | | | | | | | | | | | |  
Db 254 TTAGGAGACCACCTTCAACACCTCATCTCCAAGTAACACTACCGGCTTATCCCCAAGAG 313  
| | | | | | | | | | | | | | | | | | | | | |  
QY 1180 CCAACTCTGAGAGATCTTTTGTGTTAAACCGAANAATTACTTTTGAAAATGGACAG 123  
| | | | | | | | | | | | | | | | | | | | | |  
Db 314 CCAGCTCTGAGAAATCTTTTGATTTGAACGCGGAGGGAACACTTTTGAAAATGGACAG 373  
| | | | | | | | | | | | | | | | | | | | | |  
QY 1240 ATCTTTTCATCTCTATCAGGCTGTTGTAAGGTGATCTGAAATCAGAAATATCCAAAC 129  
| | | | | | | | | | | | | | | | | | | | | |



Mon Aug 9 13:26:10 1999

us-09-049-696-19.rst

Page 12

Db 480 GTGTCATCCCGC 492

Search completed: August 6, 1999, 13:33:58  
Job time: 13036 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:59:09 ; Search time 3268.17 Seconds  
(without alignments)  
2902.806 Million cell updates/sec

Title: US-09-049-696-20

Perfect score: 2983  
Sequence: 1 GAAATCAGGAGGATGTAC.....AATGCTAACAACATGCGTA 2983

Scoring table: OLIGO\_NDC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_vl:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_to:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hg1:\*  
35: gb\_hg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2758	92.5	3311	11	AF039400	Homo sapi

2	2757	92.4	2826	42	AF127036	AF127036 Homo sapi
3	685	23.0	878	5	I95746	I95746 Sequence 8
4	436	14.6	35278	11	AF039401	AF039401 Homo sapi
5	35	1.2	2937	12	AB017156	AB017156 Mus muscu
6	22	0.7	167880	11	HU091323	HU091323 Human Chrom
7	21	0.7	94038	8	ATAC004401	AC004401 Arabidops
8	21	0.7	86686	8	ATAC004786	AC004786 Arabidops
9	21	0.7	179854	34	CEY19P2	AT021347 Caenorhab
10	21	0.7	163735	35	AC007186	AC007186 Drosophi
11	21	0.7	39973	36	CEM106	245935 Caenorhabd
12	21	0.7	191356	42	AC005041	AC005041 Homo sapi
13	20	0.7	216750	1	BSUB0007	299110 Bacillus su
14	20	0.7	2984	3	BT036445	UB56445 Bos taurus
15	20	0.7	83544	7	AB025615	AB025615 Arabidops
16	20	0.7	143819	11	AC002994	AC002994 Homo sapi
17	20	0.7	158905	11	AC005666	AC005666 Homo sapi
18	20	0.7	598	12	MM049642	U49642 Mus musculi
19	20	0.7	173929	35	AC005941	AC005941 Drosophi
20	20	0.7	209071	37	AC005285	AC005285 Drosophi
21	20	0.7	12940	37	AB001386	AB001386 Plasmodiu
22	20	0.7	126732	42	AC007114	AC007114 Homo sapi
23	20	0.7	130020	42	HDAC004525	AC004525 Homo sapi
24	19	0.6	13246	1	KPNFUL	M32613 K. pneumonia
25	19	0.6	4914	2	AF001171	AF001171 Ralstonia
26	19	0.6	16609	3	OC063091	U63091 Oryctolagus
27	19	0.6	3645	4	D83273	D83273 Cyprinus ca
28	19	0.6	2184	5	A30438	A30438 H.sapiens p
29	19	0.6	43326	7	SPCC1223	AL031579 S.pombe c
30	19	0.6	89840	8	ATAC004705	AC004705 Arabidops
31	19	0.6	90735	8	ATF28P10	AL049655 Arabidops
32	19	0.6	175825	9	HS106H8	Z97195 Human DNA s
33	19	0.6	135686	9	HS473J5	Z83827 Human DNA s
34	19	0.6	128169	9	HS849L7	AL008987 Human DNA
35	19	0.6	2064	9	HUMIFNRG	J03143 Human inter
36	19	0.6	180388	9	HUMREBLAS	L11910 Human retin
37	19	0.6	135928	10	HS434014	AL022398 Homo sapi
38	19	0.6	91201	10	HSDA86H19	AL049562 Human DNA
39	19	0.6	125787	10	HS13817A	Z72519 Human DNA s
40	19	0.6	110384	11	AC000117	AC000117 Homo sapi
41	19	0.6	173766	11	AC004049	AC004049 Homo sapi
42	19	0.6	161326	11	AC004067	AC004067 Homo sapi
43	19	0.6	83549	11	AC004752	AC004752 Homo sapi
44	19	0.6	124191	11	AC005153	AC005153 Homo sapi
45	19	0.6	101110	37	AC005555	AC005555 Drosophi

## ALIGNMENTS

RESULT 1	AF039400	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	TITLE	REFERENCE	AUTHORS	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	
AF039400	AF039400	3311 bp	mRNA	AF039400	AF039400	AF039400.1	GI:4009457	human.	Homo sapiens	human.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	1 (bases 1 to 3311)	Gruber,A.D., Eible,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.	Genomic cloning, molecular characterization, and functional analysis of human CICAL, the first human member of the family of Ca2+-activated Cl- channel proteins	Genomics 54 (2), 200-214 (1998)	2 (bases 1 to 3311)	Gruber,A.D., Eible,R. and Pauli,B.U.	Direct Submission

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

## FEATURES

## Source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue-type="small intestine"  
1. .3311  
/gene="hclca1"  
352. .3096  
/gene="hclca1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"

## gene

## CDS

/translation="MGPKSSVFILILHLELEGALNSLIQLNNNGEIVLIDPNVP  
EDETLIQIKDMVTOASLYLFETGRKRFKYNVALIPETWTKADYVPRKLETYKNA  
DVLVAESTPGNDPEYEQMGNGEKERIHLPDI IACKLAIEYPOGKAHYEWAH  
LRMGVDEYNNDKEFYLSNGRIQAVRSAGIIGTVYKKCGGSCYTKRCKTSTKNTVL  
YKSGFEVLOSROTEKASIMFOHDSIVEFCEQONHNEKPKOMKONKLSMEVTL  
RSEDEPKTTPMTTOPNPNTPSLQIGQATVCIIVLKSSSMATGNKLNKLNAGQLEL  
LQTVELSGWGVATFDAAHVOSELQINSGRDRLARLPAAAGGTSISGKLSA  
FVIRKRYPTDSEIVLTDGEDNTISGCFNEKQSGAIHTVALSPASQLEBELSK  
MTGQLQTVASDVONNGLIIDAFGALSNGAVSQRISQLESGLILONQMANGVIV  
DSIVGKDTLFLIMTTPQIILMDSSGNGAVSQRISQLESGLILONQMANGVIV  
SLOASQTLTLTSTRASNAITLPITVTSKNTDCKDYSPTVYVANNIQQSGALPRA  
SVTALIESVNGKVTLEILLDNAGADATPDGQVSYPTVYVANNIQQSGALPRA  
AARRRYIPOOSGALYIPGWIENDEIOMNPRPINKDVIQHKOVCSRSISGSGSVAN  
DVPMATPDLFPPGQITLDAETHEGSLINLTWTAAGDVIQHKOVCSRSISGSGSVAN  
LBDKNEISQVNTVALIPKANESEVFLPENITFENGDTDLFIAIQAVDKYDLKSEI  
SNIRVSLFIPQTPETPSPDTSAPCNHINISIPGIIHILKIMWKVIGLQLSIA"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 92.5%; Score 2758; DB 11; Length 3311;  
Best Local Similarity 99.9%; Pred. No. 0;

Matches 2978; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 3 AATCAAGAGGAGATGTACACATGGGGCCATTAGAGCTTCTGTGATCTTGAATCT 62  
DB 330 AATCAAGAGGAGATGTACACATGGGGCCATTAGAGCTTCTGTGATCTTGAATCT 389  
QY 63 TCACCTTGAAGAGGGCCCTGAGTAATTCATCTACCTGACCTGACCAACATGGCTATGA 122  
DB 390 TCACCTTGAAGAGGGCCCTGAGTAATTCATCTACCTGACCTGACCAACATGGCTATGA 449  
QY 123 AGGCATTGCTGTGCAATGCACCCCAATGTGCAGAAAGTGAACACATCAATCAACAAAT 182  
DB 450 AGGCATTGCTGTGCAATGCACCCCAATGTGCAGAAAGTGAACACATCAATCAACAAAT 509  
QY 183 AAGGACATGTGTGACCCAGGATCTGTATCTGTTTGAAGCTACAGGAAAGGATTTTA 242  
DB 510 AAGGACATGTGTGACCCAGGATCTGTATCTGTTTGAAGCTACAGGAAAGGATTTTA 569  
QY 243 TTTCAAAATATGTGCAATTTGATCTGTAACATGGAAGACAAAGGCTGATATGTAG 302  
DB 570 TTTCAAAATATGTGCAATTTGATCTGTAACATGGAAGACAAAGGCTGATATGTAG 629  
QY 303 ACCAAAAGTGAACCTACAAAAATCTGATGTCTGCTGAGTCTACTCTCTCAGG 362  
DB 630 ACCAAAAGTGAACCTACAAAAATCTGATGTCTGCTGAGTCTACTCTCTCAGG 689  
QY 363 TATATGTAACCTTACATGAGACATGGCAACTGTGAGAGAAAGGTGAAGATCCA 422  
DB 690 TATATGTAACCTTACATGAGACATGGCAACTGTGAGAGAAAGGTGAAGATCCA 749  
QY 422 CCTCACTCTGATTTTCAATGACAGAAAAAGTTAGTGAATATGACCAAGAGTAGGCG 482

DB 750 CCTCACTCTGATTTTCAATGACAGAAAAAGTTAGTGAATATGACCAAGAGTAGGCG 809  
QY 483 ATTTGTCATGATGGGCTCATCTACGATGGGAGATTTGACGATTAATATGATGA 542  
DB 810 ATTTGTCATGATGGGCTCATCTACGATGGGAGATTTGACGATTAATATGATGA 869  
QY 543 GAAATCTACTATTCATCAATGAGAAATACAGCAGTAAGATGTCAGCAGTATTACTGG 602  
DB 870 GAAATCTACTATTCATCAATGAGAAATACAGCAGTAAGATGTCAGCAGTATTACTGG 929  
QY 603 TACAATGTATGTAAGAAATGTGAGGAGCGAGCTGTTACACCAAAAAGATGCACATTCA 662  
DB 930 TACAATGTATGTAAGAAATGTGAGGAGCGAGCTGTTACACCAAAAAGATGCACATTCA 989  
QY 663 TAAAGTACAGGACTATGAAAAAGATGATGTTGTTCCCAATCCCGCAGACGGA 722  
DB 990 TAAAGTACAGGACTATGAAAAAGATGATGTTGTTCCCAATCCCGCAGACGGA 1049  
QY 723 GAAGCTTCTATTAATGTTTGACACACATGTTGATCTATGTTGAATTTCTGTACAGACA 782  
DB 1050 GAAGCTTCTATTAATGTTTGACACACATGTTGATCTATGTTGAATTTCTGTACAGACA 1109  
QY 783 AAACCAAAAGAAAGCTTCCAAACAGCAAAATCAAAATCCAAATCTCCGAAGCACATG 842  
DB 1110 AAACCAAAAGAAAGCTTCCAAACAGCAAAATCAAAATCCAAATCTCCGAAGCACATG 1169  
QY 843 GGAAGTACCTGATATTCGAGAGACTTTAGAAACCACTCTATGACAAACAGCCACC 902  
DB 1170 GGAAGTACCTGATATTCGAGAGACTTTAGAAACCACTCTATGACAAACAGCCACC 1229  
QY 903 AAATCCACCTTCTCATGCTGCTGCAATGAGACAAAGATGTTGTTAGTCTTGACAA 962  
DB 1230 AAATCCACCTTCTCATGCTGCTGCAATGAGACAAAGATGTTGTTAGTCTTGACAA 1289  
QY 963 ATCTGGAACGATGGGAGCTGTAACCGGCTCAATGAGTGAATCAAGAGGCGACCTTT 1022  
DB 1290 ATCTGGAACGATGGGAGCTGTAACCGGCTCAATGAGTGAATCAAGAGGCGACCTTT 1349  
QY 1023 CCTGCTGACAGAGTGTGAGCTGGGCTGCTGGGTTGGATGTGATTTGACAGTGTGC 1082  
DB 1350 CCTGCTGACAGAGTGTGAGCTGGGCTGCTGGGTTGGATGTGATTTGACAGTGTGC 1409  
QY 1083 CCATGTACAAAGTGAATCATACATTAACAGTGGCAGTGCAGGACACACTGCGCAA 1142  
DB 1410 CCATGTACAAAGTGAATCATACATTAACAGTGGCAGTGCAGGACACACTGCGCAA 1469  
QY 1143 AAGATTACCTGACAGAGCTTCAAGAGGAGCGTCATGCAAGCGGCGTTCGATGGCATT 1202  
DB 1470 AAGATTACCTGACAGAGCTTCAAGAGGAGCGTCATGCAAGCGGCGTTCGATGGCATT 1529  
QY 1203 TACTGTGATTAGGAAGAAATATCCAACTGATGATCTGAATTTGCTGCTGACGAGATG 1262  
DB 1530 TACTGTGATTAGGAAGAAATATCCAACTGATGATCTGAATTTGCTGCTGACGAGATG 1589  
QY 1263 GGAAGACAACTATTAAGTGGGTCTTTAAGAGAGTCAAAACAAAGTGGTCCATCATCA 1322  
DB 1590 GGAAGACAACTATTAAGTGGGTCTTTAAGAGAGTCAAAACAAAGTGGTCCATCATCA 1649  
QY 1323 CACAGTGGCTTGGGGCCCTGCGAGCTCAAGAACTAAGAGGCTGTCCAAAATGACAGG 1382  
DB 1650 CACAGTGGCTTGGGGCCCTGCGAGCTCAAGAACTAAGAGGCTGTCCAAAATGACAGG 1709  
QY 1383 AGTTTACAGACATATGCTTCAGATCAAGTTCAAGAACTAGGCTCATATGATGCTTTGG 1442  
DB 1710 AGTTTACAGACATATGCTTCAGATCAAGTTCAAGAACTAGGCTCATATGATGCTTTGG 1769  
QY 1443 GGCCCTTTCATCAGGAATGAGAGCTGTCTCAGAGCTTCATCCAGCTTGAAGATGAAGG 1502  
DB 1770 GGCCCTTTCATCAGGAATGAGAGCTGTCTCAGAGCTTCATCCAGCTTGAAGATGAAGG 1829  
QY 1503 ATTAACCTCCAGAAACAGCCAGTGGATGATGACAGTGTGCGGACACACGCTGGG 1562

Db 1830 ATTAACCTCCAAACAGCCAGTGGATGAATGGCAGATGCTGGACAGACCCGTGG 1889  
QY 1563 AAGGACACTTTGTTTCTTATACCTGGACAGCAGCCTCCCCAAATCCTTCTGGGA 1622  
Db 1890 AAGGACACTTTGTTTCTTATACCTGGACAGCAGCCTCCCCAAATCCTTCTGGGA 1949  
QY 1623 TCCCACTGGACAGAGAGAGTGGCTTTGTAGTGGACAAAAACCAAAATGGCTTACT 1682  
Db 1950 TCCCACTGGACAGAGAGAGTGGCTTTGTAGTGGACAAAAACCAAAATGGCTTACT 2009  
QY 1683 CCAATCCAGGACTTCTAGAGTGGACATTTGGAATATAGTGGACAGCACTGACA 1742  
Db 2010 CCAATCCAGGACTTCTAGAGTGGACATTTGGAATATAGTGGACAGCACTGACA 2069  
QY 1743 AACCTGACCTGACTGCTCAGCTCCGCTGGCTCAATGCTACCTGCTCCAAATTAAGT 1802  
Db 2070 AACCTGACCTGACTGCTCAGCTCCGCTGGCTCAATGCTACCTGCTCCAAATTAAGT 2129  
QY 1803 GACTTCCAAACAGACAGACAGACCAAAATTCGCCAGCCCTGTGTATTTGAAA 1862  
Db 2130 GACTTCCAAACAGACAGACAGACCAAAATTCGCCAGCCCTGTGTATTTGAAA 2189  
QY 1863 TATTGCGAAGAGAGCTCCCAATTCAGAGCCAGTGTACAGACCTGATTTGAACTAGT 1922  
Db 2190 TATTGCGAAGAGAGCTCCCAATTCAGAGCCAGTGTACAGACCTGATTTGAACTAGT 2249  
QY 1923 GAATGAAAAAAGAGTACCTTGGAACTACTGGATATGAGCAGAGTGTGCTGACTAA 1982  
Db 2250 GAATGAAAAAAGAGTACCTTGGAACTACTGGATATGAGCAGAGTGTGCTGACTAA 2309  
QY 1983 GGATGAGGCTGTCTACTCAAGGTTATTTACAACTATGACAGCAATGTGTATGACTGT 2042  
Db 2310 GGATGAGGCTGTCTACTCAAGGTTATTTACAACTATGACAGCAATGTGTATGACTGT 2369  
QY 2043 AAAAGTGGGGCTCTGGGAGAGTTAAGCAGCAGAGAGAGTGTATACCCAGCAGAG 2102  
Db 2370 AAAAGTGGGGCTCTGGGAGAGTTAAGCAGCAGAGAGAGTGTATACCCAGCAGAG 2449  
QY 2103 TGGAGCAGCTGTACATACCTGGCTGGATGAGATGATGAATATACATGAAATCCAGCAG 2162  
Db 2430 TGGAGCAGCTGTACATACCTGGCTGGATGAGATGATGAATATACATGAAATCCAGCAG 2489  
QY 2163 ACCTGAAATTAATAGATATGTTCAACACAGCAAGTGTGTTCAGCAGAAATCTCTC 2222  
Db 2490 ACCTGAAATTAATAGATATGTTCAACACAGCAAGTGTGTTCAGCAGAAATCTCTC 2549  
QY 2223 GGGAGGCTCTTTGTGGCTGTGATGTCCTCAAAATGCTCCCAATCCTGATCTTCCACC 2282  
Db 2550 GGGAGGCTCTTTGTGGCTGTGATGTCCTCAAAATGCTCCCAATCCTGATCTTCCACC 2609  
QY 2283 TGGCCAATCACCGACCTGAGAGGGGGAATTCACGGGGGAGTCTCATTAATCTGACTTG 2342  
Db 2610 TGGCCAATCACCGACCTGAGAGGGGGAATTCACGGGGGAGTCTCATTAATCTGACTTG 2669  
QY 2343 GACAGCTCTCTGGGATATATGACCATGGAACGCTCACAAGTATATCATTCGAATAG 2402  
Db 2670 GACAGCTCTCTGGGATATATGACCATGGAACGCTCACAAGTATATCATTCGAATAG 2729  
QY 2403 TACAAGATTTCTGATGTCAGAGCAAGTCAATGAATCTCTCAAGTGAATACTACTGC 2462  
Db 2730 TACAAGATTTCTGATGTCAGAGCAAGTCAATGAATCTCTCAAGTGAATACTACTGC 2789  
QY 2463 TCTCATCCCAAGAGAGCACTCTGAGAGAGTCTTTTGTATTAACAGAGAAACATTAAC 2522  
Db 2790 TCTCATCCCAAGAGAGCACTCTGAGAGAGTCTTTTGTATTAACAGAGAAACATTAAC 2849  
QY 2523 TTTTGAAGATGCGACAGATCTTTTCAATGCTATTCAAGCTGTGTATTAAGTGCATCTGAA 2582  
Db 2850 TTTTGAAGATGCGACAGATCTTTTCAATGCTATTCAAGCTGTGTATTAAGTGCATCTGAA 2909  
QY 2583 ATGAGAAATTCAGAACTTGCAGAGATCTTTGTTATTCCTCCACAGACTCCGCCAGA 2642  
Db 2910 ATGAGAAATTCAGAACTTGCAGAGATCTTTGTTATTCCTCCACAGACTCCGCCAGA 2969

QY 2643 GACACCTAGTCTCGATGAAACGCTGCTCTTCTCTAATATTCATATCAACAGCAGAT 2702  
Db 2970 GACACCTAGTCTCGATGAAACGCTGCTCTTCTCTAATATTCATATCAACAGCAGAT 3029  
QY 2703 TCTGGCATTCACATTTTAAATTTATGTGGAAGTGTAGAGAACGAGCTGATCAAT 2762  
Db 3030 TCTGGCATTCACATTTTAAATTTATGTGGAAGTGTAGAGAACGAGCTGATCAAT 3089  
QY 2763 AGCCTAGGGCTGAATTTTGTCTCAGATTAATTAATTAATCAATTCATCTTTTGTGAT 2822  
Db 3090 AGCCTAGGGCTGAATTTTGTCTCAGATTAATTAATTAATCAATTCATCTTTTGTGAT 3148  
QY 2823 ATTAATTTTCTAAATGTTATTTAGACTTCTCTGAGGGGCGATATCTAAATGTTAT 2882  
Db 3149 ATTAATTTTCTAAATGTTATTTAGACTTCTCTGAGGGGCGATATCTAAATGTTAT 3308  
QY 2883 AGTACATTTTACTTAATGTTATCTGTAGGGGCGATATCTAAATGTTATTTAGACT 2942  
Db 3209 AGTACATTTTACTTAATGTTATCTGTAGGGGCGATATCTAAATGTTATTTAGACT 3268  
QY 2943 CCTGTAGGGGCGATTAATTAATGCTAAACACTGGGTA 2983  
Db 3269 CCTGTAGGGGCGATTAATTAATGCTAAACACTGGGTA 3309

RESULT 2  
AF127036  
LOCUS AF127036 2826 bp mRNA PRI 16-APR-1999  
DEFINITION Homo sapiens calcium-activated chloride channel protein 1 (CaCC1)  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2826)  
AUTHORS Agnel,M., Vernat,T. and Culouscou,J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel,M. and Culouscou,J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Neuilly-Malmaison 92500, France

FEATURES  
source  
1..2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/issue\_type="small intestine; colon"  
1..2826  
/gene="CaCC1"  
5..2749  
/gene="CaCC1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PIR:4585469"  
/translation="MGPEKSVFIIILHLLEGALSLSLIOLNNNGEIVAIIDPNV  
EDETLQIKDMVQASLYLFRATGRKREFEKVAIILIPETWTKTADYVAPKLETKNA  
DVLVASTPFGNDPEPTEDMGNCGERGEIHLTPPIAKKIABEGPOBRATVHEKHA  
LKMGVDEFYNDDEKFTLSNGRIQAVKCSAGITGVNKKCOGGSCTKRCITNKVTGL  
YKSGEFYVLSQRTKASIMFQHVDSIVFCTEDQHNKNEAPKONKCNLSWEVI  
RDSDEFKTPPTPTOPNPPTFSLIDIGRIYGLVDKSSMATGRNLNLAGOLF  
LQTVELGSSVWGVTFDSAAHVQSELIQINSQSDRPTLAKRPLAAAGSGTSGLSRA  
FVYIKKRYPTDSSEIVLLDGEDNTISCFENKVGSGATIHVVALGSPSAQDLEFLSK  
MTGGIQTIASDOYQNNGLIDAFGLSSGNGAVSQRSIQLESGLTLONSQMNNGIVIV

DSVCKDITLFLITHTTQPPQIILMDPSGKOGFVVDKTKAYLIDIPGIATVETWKY  
SIQASSQTLIVTSTRASNATLEPITVTSKTNKDSKEPSPLVIVANINOGASPILRA  
SVYALIESVNGKIVLELDLNGADATDDVSYEFYTYNGVNGSYKVALGGLVNA  
AARRVYIPDOOSGALYIPGWLENDIOWNPREDIDVQHOVCPSRSSGSEFVAS  
DYDNAPILPDLFPGQITDLEKAEIHGSLNLMETADGDYDHTAKRYIIRISTSID  
LRKNEISLOVNTTALIPREANSEVEFLRKPENITPENETDLEFIAOANDKDIJSEI  
SNARVSLFIPPOTPPETSPBDETSAPCNHINHSITPBIHILIKIMWKIGELQJLSIA

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 92.4% Score 2757; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2807; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 21 AGCAATGGGGCATTAGAGATTCTGTTCATCTTATTCCTTACCTTCTAGAGGGC 80  
DB 1 AGCAATGGGGCATTAGAGATTCTGTTCATCTTATTCCTTACCTTCTAGAGGGC 60  
QY 81 CCTGAGTATTCACCTATTCAGCTGAACAACAATGGCTATGAAGGCATTGTGCAAT 140  
DB 61 CCTGAGTATTCACCTATTCAGCTGAACAACAATGGCTATGAAGGCATTGTGCAAT 120  
QY 141 CGACCCCAATGTGCGAAGATGAACACTTCATCAACATAAAGAGCATGGTGACCA 200  
DB 121 CGACCCCAATGTGCGAAGATGAACACTTCATCAACATAAAGAGCATGGTGACCA 180  
QY 201 GGCATCTCTGTATCTGTGTAAGCTACAGAAAGCGATTTTATTTCAAAATGTGCCAT 260  
DB 181 GGCATCTCTGTATCTGTGTAAGCTACAGAAAGCGATTTTATTTCAAAATGTGCCAT 240  
QY 261 TTGTGATTCCTGAACAATGGAAGCAAGGCTGCTATGTGAGCCAAACTTGAACCTA 300  
DB 241 TTGTGATTCCTGAACAATGGAAGCAAGGCTGCTATGTGAGCCAAACTTGAACCTA 300  
QY 321 CAAAATGCTGATGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 380  
DB 301 CAAAATGCTGATGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
QY 381 TGAGCAGATGGGCACTGTGAGAGAGAGGCTGAAGAGATCCACTGCTGCTGCTGCTGCT 440  
DB 361 TGAGCAGATGGGCACTGTGAGAGAGAGGCTGAAGAGATCCACTGCTGCTGCTGCTGCT 420  
QY 441 TGAGAGAAAAAAGTATGCTGAATATGACACACAAGGTAGGGCATTTGTCCATGAGGGC 500  
DB 421 TGAGAGAAAAAAGTATGCTGAATATGACACACAAGGTAGGGCATTTGTCCATGAGGGC 480  
QY 501 TCATCTACGATGGGAGATATTTGACAGATACATAATGATGAGAAATTTCTATTATCAA 560  
DB 481 TCATCTACGATGGGAGATATTTGACAGATACATAATGATGAGAAATTTCTATTATCAA 540  
QY 561 TGGAAGAAATCAACAGCTAAGATGTTCCAGAGATATTCTGCTCAATGTAGTAAAGAA 620  
DB 541 TGGAAGAAATCAACAGCTAAGATGTTCCAGAGATATTCTGCTCAATGTAGTAAAGAA 600  
QY 621 GGTGCGGGGAGGAGTGTATACCAAAAGATGCACATTCANAAAGTAAACAGGATCTA 680  
DB 601 GGTGCGGGGAGGAGTGTATACCAAAAGATGCACATTCANAAAGTAAACAGGATCTA 660  
QY 681 TGAAGAAAGATGTGATTTGTTCTCCAAATCCGCGAGAGAGAGGCTTCTATATATTT 740  
DB 661 TGAAGAAAGATGTGATTTGTTCTCCAAATCCGCGAGAGAGAGGCTTCTATATATTT 720  
QY 741 TGCAACAACATGTTGATTTGATAGTTGATTTCTGTACAGAAACCAACACAAAGAGC 800  
DB 721 TGCAACAACATGTTGATTTGATAGTTGATTTCTGTACAGAAACCAACACAAAGAGC 780  
QY 801 TCCAAACAAGCAAAATCAAAATGCAATCCGAGACCAATGGAGTATCCGATATTC 860  
DB 781 TCCAAACAAGCAAAATCAAAATGCAATCCGAGACCAATGGAGTATCCGATATTC 840  
QY 861 TGAGAGTTTAAAGAAACCACTCTATGACAAACAGCACAACAAATCCACCTTCTCAT 920

DB 841 TGAGAGCTTTAAGAAACCACTCTATGACAAACAGCCACCACCAATCCACCTTCTCAT 900  
QY 921 GCGGCGATTGGACAAAGATGTGCTTACTGCTTGCAAAATGGAGATGGCGC 980  
DB 901 GCGGCGATTGGACAAAGATGTGCTTACTGCTTGCAAAATGGAGATGGCGC 960  
QY 981 TGGTAAACCGCTCAATCGACTGAATCAAGCAGCCAGCTTTCTGCTGCACAGATTGA 1040  
DB 961 TGGTAAACCGCTCAATCGACTGAATCAAGCAGCCAGCTTTCTGCTGCACAGATTGA 1020  
QY 1041 GCTGGGCTCTGGGTTGGATGATGACATTTACAGTGTGCTGCCATGTACAAAGTGAAT 1100  
DB 1021 GCTGGGCTCTGGGTTGGATGATGACATTTACAGTGTGCTGCCATGTACAAAGTGAAT 1080  
QY 1101 CATACAGATTAACAGTGGCAGTGCAGAGGACACACTGCGCAAAAGATTACGCGCAGC 1160  
DB 1081 CATACAGATTAACAGTGGCAGTGCAGAGGACACACTGCGCAAAAGATTACGCGCAGC 1140  
QY 1161 TTCAGAGGAGACCTCAGCTCAGCGGGCTTCGATCGGCATTTACTGTATAGAGAA 1220  
DB 1141 TTCAGAGGAGACCTCAGCTCAGCGGGCTTCGATCGGCATTTACTGTATAGAGAA 1200  
QY 1221 ATATCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1280  
DB 1201 ATATCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
QY 1281 TGGGCTCTTTAAGAGGCAACAAAGTGTGCTCATTCACACAGCTGCTTGGGGC 1340  
DB 1261 TGGGCTCTTTAAGAGGCAACAAAGTGTGCTCATTCACACAGCTGCTTGGGGC 1320  
QY 1341 CTCTGCACTCAAGAACTAGAGAGAGCTGCCAAATGACAGAGGTTTACAGACATATGC 1400  
DB 1321 CTCTGCACTCAAGAACTAGAGAGAGCTGCCAAATGACAGAGGTTTACAGACATATGC 1380  
QY 1401 TTCAGATCAAGTTCAGAAACAAATGGCCTCATGATGATGATGATGATGATGATGATGAT 1460  
DB 1381 TTCAGATCAAGTTCAGAAACAAATGGCCTCATGATGATGATGATGATGATGATGATGAT 1440  
QY 1461 TGGAGCTGTCTCTCAGCGCTCCATCAGCTTGAGAGTAAAGGATTTAACCTCCAGAACG 1520  
DB 1441 TGGAGCTGTCTCTCAGCGCTCCATCAGCTTGAGAGTAAAGGATTTAACCTCCAGAACG 1500  
QY 1521 CCAGTGAATGAATGGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1580  
DB 1501 CCAGTGAATGAATGGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
QY 1581 TATCACTGAGCAACGAGCCTCCCAATCTTCTGCGGATCCAGTGGACAGAAACA 1640  
DB 1561 TATCACTGAGCAACGAGCCTCCCAATCTTCTGCGGATCCAGTGGACAGAAACA 1620  
QY 1641 AGGTGCTTTGTAGTGAACAAAACACCAAAATGGCTTACTCCAAATCCAGGCAATTGC 1700  
DB 1621 AGGTGCTTTGTAGTGAACAAAACACCAAAATGGCTTACTCCAAATCCAGGCAATTGC 1680  
QY 1701 TAAAGTGGCACTTGGAAATCAGTCTGCAAGCAAGCTCAAAACCTTGACCTGACCTGT 1760  
DB 1681 TAAAGTGGCACTTGGAAATCAGTCTGCAAGCAAGCTCAAAACCTTGACCTGACCTGT 1740  
QY 1761 CACGTCCCGTGGCTCAATGCTTACCTGCTCTCAATTCAGTGTGATTCACAAAGCAACA 1820  
DB 1741 CACGTCCCGTGGCTCAATGCTTACCTGCTCTCAATTCAGTGTGATTCACAAAGCAACA 1800  
QY 1821 GGACACAGCAAAATCCCAAGCCTCTGTGATTTATGCAAAATTTGCGCAAGAGCCTC 1880  
DB 1801 GGACACAGCAAAATCCCAAGCCTCTGTGATTTATGCAAAATTTGCGCAAGAGCCTC 1860  
QY 1881 CCGAATCTCAGGGGCGAGTGCACAGCCCTGATGATGATGATGATGATGATGATGATGAT 1940  
DB 1861 CCGAATCTCAGGGGCGAGTGCACAGCCCTGATGATGATGATGATGATGATGATGATGAT 1920  
QY 1941 CTTGAACTACTGATATGAGCAGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2000



```

REFERENCE Eutheria: Primates: Catarrhini: Homiidae: Homo.
AUTHORS 1 (bases 1 to 35278)
          Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and
          Pauli,B.U.
TITLE      Genomic cloning, molecular characterization, and functional
          analysis of human hCLCA1, the first human member of the family of
          Ca2+-activated Cl- channel proteins
JOURNAL    Genomics 54 (2), 200-214 (1998)
MEDLINE    99047526
REFERENCE  2 (bases 1 to 35278)
          Gruber,A.D., Elble,R. and Pauli,B.U.
TITLE      Direct Submission
AUTHORS    Submitted (19-Dec-1997) Department of Pathology, College of
JOURNAL    Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
FEATURES
SOURCE     1. 35278
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /chromosome="1"
           /map="1p22-p31"
           1. 1617
           /gene="hCLCA1"
           1. 33522
           /gene="hCLCA1"
           767. 1265
           /rpt_family="L1 transposable element"
           /rpt_type-dispersed
           1587. 1596
           /gene="hCLCA1"
           /note="putative"
           1618. 1708
           /gene="hCLCA1"
           /number=1
           join(1618. 1708,1962. 2383,6667. 6807,6979. 7126,6986.
           9696. 9801,15454. 15631,18591. 18809,19783. 20010,
           22253. 22427,24522. 24628,26644. 26859,27447. 27708,
           28766. 28936,31840. 32079,32919. 33522)
           /gene="hCLCA1"
           /product="calcium-dependent chloride channel-1"
           1962. 2383
           /gene="hCLCA1"
           /number=2
           join(2222. 2383,6667. 6807,6979. 7126,6986. 9801,
           15454. 15631,18591. 18809,19783. 20010,22253. 22427,
           24522. 24628,26644. 26859,27447. 27708,28766. 28936,
           31840. 32079,32919. 33310)
           /gene="hCLCA1"
           /codon_start=1
           /product="calcium-dependent chloride channel-1"
           /protein_id="AAC5429.1"
           /db_xref="PIR:g4009460"
           /db_xref="GI:4009460"
           /translation="MGPFKSFVFLLIHLLLEGALNSLIQLNNNGEIGVIAIDPNVP
           EDTLLIQIDKMTVOASILYLFKATGKRFYKKNVAILPETKTRADYVGRPLETYKNAAD
           DVIAEETPPGNDPEYEMONGCNGEKGERHILPDPFAGKLAIEGPGAGAFEMHMA
           LRNGVEPEYNNDEKFEYLSNGRIQAVRCASGTTGNVYKKGQSGCYTRKCTENKVTGL
           YKCEPEYDLSROTERAATIMFAQVHSDIVERTQNNKAPKQNKCRINSEWT
           RSEDEFRKTTPTTQPPNPTFSLQIQIRIVCLVLDKSGSWATGNRLNRNLQAQGLF
           LQVEGSGWGVMTFDSAHVQSELIQINSGSDRDTGAKRLPAASGTSICGLRSEA
           FTVIRKKYPTDGEIVLITDGEDNTIGCNEVYNSQSAIHTVALIGSAQDELEISK
           MTGGLQYVADQVQNNGLIDAFGLSSGNASVORSLSIKSGITLSONMNGVITVA
           DSVAGQDLFLITWTQPOIILMDPSGQKGGVYKRNKMAVILQIGLAKGTWY
           SIQASGQTLTLTVTSRASNATLIPRYTSSKQNDTSFPPKPIYVAINIRGASPIILA
           SYVALIESVNGKTVTLELDNGAGADTKDQVYSRIFTYTDINGRISVYRAAGVNA
           AARRRYPQSGALYIPGWLLENELQNNPRPEITNKDVQHKQVCFRTSSGGSFVNAS
           DVYNAAPIDLPFPGQITDLNAEIHGSLIMLTATAPDDDHGAHAYIIRISDIID
           LRKFESLQVNTALIPKANSSEVLFKRPENTFENGIDLFAIQADVQDKLSRI
           SNIAVSLFIPTQTPPETPSPDETSACPNIHINSTIPGIHIKIMKWIGELQLSIA
           "
exon      6667. 6807
           /gene="hCLCA1"
           /number=3
           6979. 7126

```

Query Match	Best Local Similarity	Score	DB	Length
Matches 606; Conservative 0; Mismatches 1; Indels 1; Gaps 1	99.7%;	436;	DB 11;	35278;
exon	/gene="hclcal1"	/number=4	9696..3801	
exon	/gene="hclcal1"	/number=5	15454..15631	
exon	/gene="hclcal1"	/number=6	18591..18809	
exon	/gene="hclcal1"	/number=7	19783..20010	
exon	/gene="hclcal1"	/number=8	22253..22427	
exon	/gene="hclcal1"	/number=9	24522..24628	
exon	/gene="hclcal1"	/number=10	26644..26859	
exon	/gene="hclcal1"	/number=11	27447..27708	
exon	/gene="hclcal1"	/number=12	28766..28936	
exon	/gene="hclcal1"	/number=13	31840..32079	
exon	/gene="hclcal1"	/number=14	32919..33522	
exon	/gene="hclcal1"	/number=15		
BASE COUNT	11245 a	6870 c	7230 g	9933 t
ORIGIN				
Query Match	14.6%;	Score 436;	DB 11;	Length 35278;
Best Local Similarity	99.7%;	Pred. No. 3.4e-241;		
Matches 606; Conservative 0; Mismatches 1; Indels 1; Gaps 1				
exon	/gene="hclcal1"	/number=4	9696..3801	
exon	/gene="hclcal1"	/number=5	15454..15631	
exon	/gene="hclcal1"	/number=6	18591..18809	
exon	/gene="hclcal1"	/number=7	19783..20010	
exon	/gene="hclcal1"	/number=8	22253..22427	
exon	/gene="hclcal1"	/number=9	24522..24628	
exon	/gene="hclcal1"	/number=10	26644..26859	
exon	/gene="hclcal1"	/number=11	27447..27708	
exon	/gene="hclcal1"	/number=12	28766..28936	
exon	/gene="hclcal1"	/number=13	31840..32079	
exon	/gene="hclcal1"	/number=14	32919..33522	
exon	/gene="hclcal1"	/number=15		
BASE COUNT	11245 a	6870 c	7230 g	9933 t
ORIGIN				
Query Match	14.6%;	Score 436;	DB 11;	Length 35278;
Best Local Similarity	99.7%;	Pred. No. 3.4e-241;		
Matches 606; Conservative 0; Mismatches 1; Indels 1; Gaps 1				
exon	/gene="hclcal1"	/number=4	9696..3801	
exon	/gene="hclcal1"	/number=5	15454..15631	
exon	/gene="hclcal1"	/number=6	18591..18809	
exon	/gene="hclcal1"	/number=7	19783..20010	
exon	/gene="hclcal1"	/number=8	22253..22427	
exon	/gene="hclcal1"	/number=9	24522..24628	
exon	/gene="hclcal1"	/number=10	26644..26859	
exon	/gene="hclcal1"	/number=11	27447..27708	
exon	/gene="hclcal1"	/number=12	28766..28936	
exon	/gene="hclcal1"	/number=13	31840..32079	
exon	/gene="hclcal1"	/number=14	32919..33522	
exon	/gene="hclcal1"	/number=15		
BASE COUNT	11245 a	6870 c	7230 g	9933 t
ORIGIN				

DB 33337 ATAAATCATTCATCC-TTTTGTGATTATAAAATTTCTAAATGATTTAGACTTCCT 33395  
QY 2856 GTAGGGGGCGATATCTAAATGATATAGTACATTATCTAAATGATTCCTGAGGG 2915  
DB 33396 GTAGGGGGCGATATCTAAATGATATAGTACATTATCTAAATGATTCCTGAGGG 33455  
QY 2916 GCATATACATTAATGATTTAGACTTCCTGAGGGCGCATTAATAATGCTAACA 2915  
DB 33456 GCATATACATTAATGATTTAGACTTCCTGAGGGCGCATTAATAATGCTAACA 33515  
QY 2976 ACTGGGTA 2983  
DB 33516 ACTGGGTA 33523

RESULT 5  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS Mus musculus gob-5 mRNA, complete cds.  
DEFINITION AB017156  
ACCESSION AB017156  
NID 93721911  
VERSION AB017156.1 GI:3721911  
KEYWORDS GOB-5.  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.  
REFERENCE 1 (sites)  
AUTHORS Komiyama, T., Tanigawa, Y., and Hirohashi, S.  
TITLE Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice  
JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
MEDLINE 99160866  
REFERENCE 2 (bases 1 to 2937)  
AUTHORS Komiyama, T., Tanigawa, Y., and Hirohashi, S.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/Genbank databases, Tohru Komiyama, ERATO, JST, Hirohashi Cell Configuration Project; 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan  
(E-mail:tkom@hccp.jst.go.jp, Tel:81-298-47-7563, Fax:81-298-47-5226)  
FEATURES  
source location/Qualifiers  
1. 2937  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15..2756  
/gene="gob-5"  
15..2756  
/gene="gob-5"  
/codon\_start=1  
/protein\_id="BAA33743.1"  
/db\_xref="PID:d1034712"  
/db\_xref="PID:g3721912"  
/db\_xref="GI:3721912"  
translation="MESKSPVFLILHLLEGLVLSLQLNNGNGEIVAIIDHNP  
DEBALIHKIDWYQASPYLFEATGKRFYKNVAILIPESMKRPYTPKLEFENA  
DVLSTSPLDGDEPTEHIGAGEKGRILHLPDLAGKLLQYODPRTFHEVAH  
FRMGVENEYNDKEKFLSKGKPOAVRCSAATKNGVRDGGSCITNKCVDRTG  
LYKNCVFPYDPDRHONKASIMPNQNTNSVETCKENHONAPNDONONKLRSTVEY  
IOESDFKQTPPTAOPAPPTSLDIGRIVCLVLDKSGSLNDRLNRMNASLFL  
LQATVEQSGWAGVTDSAAVQSELKOLNSGADRLILHLPYVAGSAGTISCSGLT  
AFTVIRKKYPTDGESEIVLTDGEDNTISSCFDLYKQSGAIHTVVALGPAALAELEDS  
KMTGKQTSDDOVONNGVDAFAALSGMAIAHSGIOESGVNLONNOMNGSVI  
VSSVSKDILFLITHTPTLTIIMPDSGEONGFLDITTKYAYLOVPGTAKVGFVK  
YSLQASSQTLTLTVSRASATLPTITVYPPYKNTGKRPSPYTVASIRQASPLR  
ASVTALLESNGKTVLLELDNGAGDARKNDGVISRTFAPDANRISVKIHALGAV  
TSDROAAPPKNRNAMYIDGWIEDGVRNMPRETSYVODKOLCFSTSSGSEFVAVN  
VPAAPILPLFPCCQITDLKASIOGONLVNLWTAPGDYDHGRASNYIIRSTSIVD  
LBDHFNISLOVNTVGLIPKASSEIEFEELGNTFGNGDIFIAIOAVDKSLKSEI  
SNIRSVIFIPAOEPIPEDSTPCPDISINSTRIPGIIHVLKIMMKLGMQVTLGLH"

BASE COUNT 860 a 718 c 693 g 666 t  
ORIGIN  
Query Match 1.2%; Score 35; DB 12; Length 2937;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1222 TATCCAACTGATGAGTGAATGATGCTGCTGAC 1256  
DB 1215 TATCCAACTGATGAGTGAATGATGCTGCTGAC 1249

RESULT 6  
HU091323 167880 bp DNA PRI 21-JAN-1999  
LOCUS Human Chromosome 16 BAC clone C19878K-A-972D3, complete sequence.  
DEFINITION HU091323  
ACCESSION U91323  
NID 93582311  
VERSION U91323.1 GI:3582311  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 167880)  
AUTHORS Adams, M.D., Loftus, B.J., Phillips, C.A., Zhou, L., Brandon, R. and Venter, J.C.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAR-1997) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA  
REFERENCE 2 (bases 1 to 167880)  
AUTHORS Adams, M.D., Loftus, B.J., Zhou, L., La Bombard, M., Kim, U.J. and Venter, J.C.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA  
REFERENCE 3 (bases 1 to 167880)  
AUTHORS Adams, M.D., Loftus, B.J., Zhou, L., La Bombard, M., Kim, U.J. and Venter, J.C.  
TITLE Direct Submission  
JOURNAL Submitted (22-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
REFERENCE 4 (bases 1 to 167880)  
AUTHORS Adams, M.D. and Loftus, B.J.  
TITLE Direct Submission  
JOURNAL Submitted (11-SEP-1998) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
COMMENT On Sep 11, 1998 this sequence version replaced gi:2335056.  
BAC clone C19878K-972D3 is located in band 16p13.11 of chromosome 16. Genes were identified by a combination of five methods: XBRIL (available by anonymous ftp from athur.epm.ornl.gov), GeneFinder (available by anonymous ftp from collin@u.washington.edu), GENSCAN (available using the e-mail server at genscan@genomic.stanford.edu), searches using the EST database at TIGR (http://www.tigr.org/tcd/hcd/hcd.html) and searches against a peptide database. Repeats were identified using RepeatMasker (Smit, A.F.A. and Green, P. unpublished, http://ftp.genome.washington.edu/rm/RepeatMasker.html).  
location/Qualifiers  
1. 167880  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16p13.11"  
/clone="A-972D3"  
68302..68901  
/gene="A-972D3.1"  
/pseudo  
/gene="A-972D3.2"  
/gene="A-972D3.2"

```

CDS
162350. .>162696
/gene="A-972D3.2"
/codon_start=1
/product="Myosin heavy chain (MHY1) (3'partial)"
/protein_id="AAC35212.1"
/db_xref="PID:91871206"
/db_xref="GI:1871206"
/translation="MAKGGLSDDEKFLFVDKNFNSPVAQADNAARLVMVPEKGG
FEASIKKEGDEVVVELVYENKRTVGKDILQKNPKFKSRVEDMELITLNRSAYI
HNLRRTVSGLIYV"
BASE COUNT      46429 a 38556 c 38892 g 43951 t      52 others
ORIGIN
Query Match      0.7%; Score 22; DB 11; Length 167880;
Best Local Similarity 100.0%; Pred No. 0.45;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 164 AAACACTCATTCACCAATATAA 185
|||||
|||||
Db 25031 AAACACTCATTCACCAATATAA 25052
|||||
|||||
RESULT 7
LOCUS ATAC004401 94038 bp DNA PLN 22-MAY-1999
DEFINITION Arabidopsis thaliana chromosome II BAC F21P24 genomic sequence,
ACCESSION AC004401
NID AC004169
VERSION AC004401.1 GI:3169169
KEYWORDS HTC.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis
1 (bases 1 to 94038)
Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
Sykes,S.M., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
Somerville,C.R. and Venter,J.C.
Arabidopsis thaliana chromosome II BAC F21P24 genomic sequence
Unpublished
2 (bases 1 to 94038)
Rounsley,S.D. and Lin,X.
Direct Submission
Submitted (12-MAR-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
3 (bases 1 to 94038)
Rounsley,S.D.
Direct Submission
Submitted (30-MAY-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
4 (bases 1 to 94038)
Lin,X.
Direct Submission
Submitted (22-MAY-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On May 30, 1998 this sequence version replaced g1:3142312.
Address all correspondence to:
Steve Rounsley
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850,
USA
e-mail: rounsley@tigr.org
BAC clone F21P24 is from Arabidopsis chromosome II and is contained
in the YAC clone C106C07.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp

```

from arthur.epm.ornl.gov). GeneFinder (Phil Green, University of Washington), GenScan (Chris Burge, <http://genome.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arjan Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

location/Qualifiers

1.94038

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db\_xref="taxon:3702"

/chromosome="II"

/map="C106C07"

/clone="F21P24"

1..22544

/note="overlap with BAC clone T20K9 (AC004786:64143..86686)."

<1..>1327

/gene="F21P24.1"

/note="T20K9.16"

join(1..92,168..326,400..571,679..977,1088..1327)

/gene="F21P24.1"

/codon\_start=3

/product="putative Ca<sup>2+</sup>-ATPase, 5' partial"

/protein\_id="AAC17813.1"

/db\_xref="PID:g3169170"

/db\_xref="GI:3169170"

/translation="ILTDGIAIEGPFREKNQEMLELPRIQVNAWSSPMOKHTLV  
KQRTAFDEVAVTGGINDAPALHEDAGLWGIGTEVAKELADVIITLDNFSVIV  
TAWKGRSVYINOKFQPOLTYNVALLYNFSSACTGSAPTLAVOILLNNMTNL  
GALKATSPNNNELMKRMPPGRGNFTIANNENITGOAYOPTITTIIOAKRSMPG  
LVASDSTLVNTLIENCVFPCOVFNESKEHEIDVFGLDNIFYVVVVGTVTFQ  
IILLEGTFASTPLTIWQFFSIFVFGLGMIAGLKRPV"

join(<1..92,168..326,400..571,679..977,1088..>1327)

/gene="F21P24.1"

1903..2036

/rpt\_family="(TAAA)n"

complement(2191..2239)

/note="exon predicted by xgrail, quality marginal\_shadowexon"

marginal\_shadowexon

complement(2295..2360)

/note="exon predicted by xgrail, quality good\_shadowexon"

2455..4493

/gene="F21P24.2"

/note="T20K9.17: similar to serine carboxypeptidase I"

/pseudo

repeat\_region

complement(3328..3385)

/rpt\_family="AR\_rich"

repeat\_region

complement(3390..3506)

/rpt\_family="(TAAA)n"

misc\_feature

4509..4538

/note="exon predicted by xgrail, quality marginal"

complement(4553..4590)

/rpt\_family="(TA)n"

repeat\_region

complement(4633..4653)

/rpt\_family="AR\_rich"

misc\_feature

complement(4830..4954)

/note="exon predicted by xgrail, quality marginal\_shadowexon"

misc\_feature

4848..4958



```

misc-feature      /note="exon predicted by xgrall, quality excellent"
complement(5440..5613)
/note="exon predicted by xgrall, quality good, shadowexon"
join(<6436..6560,6635..6780,6881..6957,7038..7160,
7555..7667,7736..7773,7979..8077,8179..8265,8358..8426,
8612..8683,8764..8885,9346..9459,9572..9617,9705..>9785)
/gene="F21P24.3"
<6436..>9785
/gene="F21P24.3"
/note="T20K9.18"
join(6436..6560,6635..6780,6881..6957,7038..7160,
7555..7667,7736..7773,7979..8077,8179..8265,8358..8426,
8612..8683,8764..8885,9346..9459,9572..9617,9705..9785)
/gene="F21P24.3"
/codon_start=1
/product="putative serine carboxypeptidase I"
/protein_id="AAC17814.1"
/db_xref="GI:3169171"
/db_xref="GI:3169171"
/translacion="METLTKLVLVLLFILNHHVSGSIVKFLPFGSGPLPELETGCI
GIEEEVQLFVYFIKSERPKEDPLMLTSGGSGSITGLFQNGPLAKSEVYNG
SVPSLVSTVSTKIANIIFLDQVAGSGSAPLIDPTDGEVRIHEFIQKMS
KHPOFSNHFYAGDSYSGMIVALVOEISGVYICNPINKGYVLAGRTIHEDDP
NRIPIPSHGMALISDELYESIRACKGNFNNDPRNTKCKLKEEFHKCDKINFEPI
LSPDCITASPDCTILPYPLISFANDSEVDAALHVNKRSIGKRECYLSKPYNKDK
SSVPIHNNVSQIRSLTISGDHVLVPLATQAWIKSLTYSITDEWRPMINDQITG
YTRYSNKMTEFAVIRKSGHAEKPKPESIFMRWINGOPL"
complement(7277..7399)
/rpt_family="AT_rich"
9116..9138
/rpt_family="AT_rich"
10479..10527
/note="exon predicted by xgrall, quality good"
join(<10590..10717,10800..10945,11012..11088,11168..11290,
11602..11704,11806..11828,11912..12013,12378..12464,
12552..12620,12727..12798,13609..13730,14190..14303,
14407..14452,14559..>14639)
/gene="F21P24.4"
<10590..>14639
/gene="F21P24.4"
/note="T20K9.19"
join(10590..10717,10800..10945,11012..11088,11168..11290,
11602..11704,11806..11828,11912..12013,12378..12464,
12552..12620,12727..12798,13609..13730,14190..14303,
14407..14452,14559..14639)
/gene="F21P24.4"
/codon_start=1
/product="putative serine carboxypeptidase I"
/protein_id="AAC17815.1"
/db_xref="PIR:G3169172"
/db_xref="GI:3169172"
/translacion="MSLTLEFLLLIVLLISHAHSGSIVKFLPFGSGPLPELETGCI
IGIEEEVQLFVYFIKSERPKEDPLMLTSGGSGSITGLFQNGPLAKSEVYNG
SVPSLVSTVSTKIANIIFLDQVAGSGSAPLIDPTDGEVRIHEFIQKMS
KHPOFSNHFYAGDSYSGMIVALVOEISGVYICNPINKGYVLAGRTIHEDDP
NRIPIPSHGMALISDELYESIRACKGNFNNDPRNTKCKLKEEFHKCDKINFEPI
LSPDCITASPDCTILPYPLISFANDSEVDAALHVNKRSIGKRECYLSKPYNKDK
SSVPIHNNVSQIRSLTISGDHVLVPLATQAWIKSLTYSITDEWRPMINDQITG
YTRYSNKMTEFAVIRKSGHAEKPKPESIFMRWINGOPL"
13918..14008
/rpt_family="(TA)n"
complement(14826..14989)
/note="exon predicted by xgrall, quality good"
complement(17227..17330)
/note="exon predicted by xgrall, quality
marginal, shadowexon"
17429..17488
/note="exon predicted by xgrall, quality marginal"
join(<17954..18072,18144..18289,18426..18502,18632..18751,
19006..19108,19205..19242,19338..19439,19871..19957,
20049..20126,20326..20397,20476..20594,21039..21152,
21262..21307,21406..>21486)
/gene="F21P24.5"

```

```

gene      <17954..>21486
          /gene="F21P24.5"
          /note="T20K9.20"
CDS      join(17954..18072,18144..18289,18426..18502,18632..18751,
19006..19108,19205..19242,19338..19439,19871..19957,
20049..20126,20326..20397,20476..20594,21039..21152,
21262..21307,21406..>21486)
          /gene="F21P24.5"

Query Match      0.7% Score 21; DB 8; Length 94038;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2798 AATCATCATCCTTTT 2818
Db 9541 AATCATCATCCTTTT 9561

RESULT 8
LOCUS ATAC004786
DEFINITION Arabidopsis thaliana chromosome II BAC T20K9 genomic sequence,
complete sequence.
ACCESSION AC004786
NID G3445196
VERSION AC004786.1 GI:3445196
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (bases 1 to 86686)
Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
Sykes,S.M., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
Somerville,C.R. and Venter,D.C.
Arabidopsis thaliana chromosome II BAC T20K9 genomic sequence
Unpublished
2 (bases 1 to 86686)
Rounsley,S.D. and Lin,X.
Direct Submission
Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, rounsley@tigr.org
3 (bases 1 to 86686)
Rounsley,S.D.
Direct Submission
Submitted (20-AUG-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
4 (bases 1 to 86686)
Lin,X.
Direct Submission
Submitted (22-MAY-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Aug 20, 1998 this sequence version replaced gi:3420041.
Address all correspondence to:
Steve Rounsley
The Institute for Genomic Research
9712 Medical Center Dr,
Rockville, MD 20850,
USA
e-mail: rounsley@tigr.org
BAC clone T20K9 is from Arabidopsis chromosome II and is contained
in the YAC clone C1C06C07.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAT (available by anonymous ftp
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://genome.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
(http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tib/at.html).
Annotated genes are named to indicate the level of evidence for

```



Query Match 0.7%; Score 21; DB 8; Length 86686;  
 Best Local Similarity 100.0%; Pred. No. 1.7;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2798 AATCATTCATCCCTTTT 2818  
 |||||||  
 Db 73683 AATCATTCATCCCTTTT 73703

RESULT 9  
 CEY19D2/c  
 LOCUS  
 DEFINITION Caenorhabditis elegans chromosome II clone Y19D2. WORKING DRAFT  
 SEQUENCE, in unordered pieces.  
 ACCESSION AL021347  
 MID 93646815  
 VERSION AL021347.1 GI:3646815  
 KEYWORDS HTG: HTGS\_PHASE1.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;  
 Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 179854)  
 AUTHORS White,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-OCT-1998) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jesse@wanger.ac.uk or rwenematode.wustl.edu  
 On Sep 24, 1998 this sequence version replaced gi:2804161.  
 IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

FEATURES  
 Location/Qualifiers  
 source 1..179854  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="II"  
 /clone="Y19D2"  
 BASE COUNT 56786 a 32609 c 32628 g 56230 t 1601 others  
 ORIGIN

Query Match 0.7%; Score 21; DB 34; Length 179854;  
 Best Local Similarity 100.0%; Pred. No. 1.7;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2826 AATTTTCTAAATGATTTT 2846  
 |||||||  
 Db 74101 AATTTTCTAAATGATTTT 74081

RESULT 10  
 AC007186/c  
 LOCUS  
 DEFINITION AC007186 163735 bp DNA HTG 05-APR-1999  
 Drosophila melanogaster chromosome 2 clone BACR03D06 (D569) RPCI-98  
 03.D.6 map 32A-32A strain Y2; cn bw sp, WORKING DRAFT SEQUENCE, 20  
 unordered pieces.  
 ACCESSION AC007186  
 MID 94559424  
 VERSION AC007186.3 GI:4559424  
 KEYWORDS HTG: HTGS\_PHASE1.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

REFERENCE  
 AUTHORS  
 1 (bases 1 to 163735)  
 Celniker,S.E., Aspayani,A., Arcana,T.T., Baxter,E., Blazej,R.G.,  
 Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Houston,K.A., Hummatti,S.R., Karra,K., Kearney,L.,  
 Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
 Pfeiffer,B., Poon,L., Sequelra,A., Sethi,H., Snir,E.,  
 Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
 Rubin,G.M.  
 Sequencing of Drosophila melanogaster  
 Unpublished  
 2 (bases 1 to 163735)  
 Celniker,S.E., Aspayani,A., Arcana,T.T., Baxter,E., Blazej,R.G.,  
 Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Houston,K.A., Hummatti,S.R., Karra,K., Kearney,L.,  
 Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
 Pfeiffer,B., Poon,L., Sequelra,A., Sethi,H., Snir,E.,  
 Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
 Rubin,G.M.  
 Direct Submission  
 Submitted (31-MAR-1999) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On Apr 5, 1999 this sequence version replaced gi:4558629.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet  
 the following cutoffs: length >= 400 bases, phrap computed error  
 rate <= 1/10.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 20 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 750: contig of 750 bp in length  
 751 830: gap of unknown length  
 831 1608: contig of 778 bp in length  
 1609 1688: gap of unknown length  
 1689 2959: contig of 1271 bp in length  
 2960 3039: gap of unknown length  
 3040 3675: contig of 636 bp in length  
 3676 3755: gap of unknown length  
 3756 4524: contig of 769 bp in length  
 4525 4604: gap of unknown length  
 4605 5221: contig of 617 bp in length  
 5222 5301: gap of unknown length  
 5302 6242: contig of 941 bp in length  
 6243 6322: gap of unknown length  
 6323 8150: contig of 1828 bp in length  
 8151 8230: gap of unknown length  
 8231 11358: contig of 3128 bp in length  
 11359 11438: gap of unknown length  
 11439 13567: contig of 2129 bp in length  
 13568 13647: gap of unknown length  
 13648 18209: contig of 4562 bp in length  
 18210 18289: gap of unknown length  
 18290 21626: contig of 3337 bp in length  
 21627 21706: gap of unknown length  
 21707 26080: contig of 4374 bp in length  
 26081 26160: gap of unknown length  
 26161 32107: contig of 5947 bp in length  
 32108 32187: gap of unknown length  
 32188 45071: contig of 12884 bp in length  
 45072 45151: gap of unknown length  
 45152 56388: contig of 11237 bp in length

\* 56389 56468: gap of unknown length  
\* 56469 75510: contig of 19042 bp in length  
\* 75511 75590: gap of unknown length  
\* 75591 102704: contig of 27114 bp in length  
\* 102705 102784: gap of unknown length  
\* 102785 129065: contig of 26281 bp in length  
\* 129066 129145: gap of unknown length  
\* 129146 163735: contig of 34590 bp in length.  
Location/Qualifiers  
1. 163735  
/organism="Drosophila melanogaster"  
/strain="v2: cn bw sp"  
/db\_xref="taxon:7227"  
/clone="BACR03D06 (D569) RPCI-98 03.D.6"  
/chromosome="2"  
/clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
PBAC3.6"  
/map="32A-32A"  
BASE COUNT 47639 a 33822 c 34032 g 46722 t 1520 others  
ORIGIN  
Query Match 0.7%: Score 21; DB 35; Length 163735;  
Best Local Similarity 100.0%: Fred. No. 1.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2121 TGGCTGATTGAGATGATGA 2141  
|||||  
Db 94913 TGGCTGATTGAGATGATGA 94893  
RESULT 11  
CEMI06/c 39973 bp DNA INV 23-NOV-1998  
LOCUS CEMI06  
DEFINITION Caenorhabditis elegans cosmid M106, complete sequence.  
ACCESSION Z46935  
KEYWORDS 246935.1 GI:602783  
VERSION HTG; conglutin: F-actin capping protein; GMP synthase; nucleotide  
binding; quinone oxidoreductase; transfer RNA; tRNA-Leu.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;  
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 39973)  
Palmer, S.  
REFERENCE Direct Submission  
AUTHORS Submitted (16-DEC-1994) Louis, MO 63110, USA. E-mail:  
jesesanger.ac.uk or rwenematode.wustl.edu  
JOURNAL 2 (bases 1 to 39973)  
2 (bases 1 to 39973)  
REFERENCE Wilson, R., Alnscough, R., Anderson, K., Baynes, C., Berks, M.,  
Bifford, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,  
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,  
Fullon, L., Gardner, A., Green, P., Hawkins, J., Hillier, L., Jier, M.,  
Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N.,  
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,  
O'Callaghan, M., Parsons, J., Percy, C., Ritten, L., Roopra, A.,  
Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,  
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaundin, M.,  
Vaughan, K., Waterston, R., Watson, A., Wellstock, L.,  
Wilkinson-Sproat, J. and Wohlman, P.  
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans  
JOURNAL Nature 368 (6466), 32-38 (1994)  
MEDLINE 94150718  
COMMENT Coding sequences below are predicted from computer analysis, using  
predictions from GeneFinder (P. Green, U. Washington), and other  
available information.  
For a graphical representation of this sequence and its analysis  
see:  
http://webace.sanger.ac.uk/cgi-  
bin/displaypdb-wormaceclass-sequence?object=M106

FEATURES  
SOURCE  
gene  
CDS  
gene  
CDS  
Location/Qualifiers  
1. 39973  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="II"  
/clone="M106"  
complement(10154..16153)  
/gene="M106.2"  
complement(join(10154..10357,10904..11026,11275..11466,  
12100..12178,13101..13253,13649..13765,13809..13946,  
14493..14716,14781..14970,16014..16153))  
/gene="M106.2"  
/note="Weak similarity to conglutin (PIR accession number  
A33090); CDNA EST EMBL:701355 comes from this gene; CDNA  
EST yk383f3.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CA87049.1"  
/db\_xref="PID:e1351703"  
/db\_xref="PID:g3878712"  
/db\_xref="GI:3878712"  
/db\_xref="SPTREMBL:009592"  
/translation="MDGPNICPGNLSVTHARKNCRKNSLKLCISIRARCLEIM  
DKRELDRIVEIKSTISLANFNQNTSFTSQGSTASTQSTSDVLRSTFRSV  
NOSTDTPPEIASFSKDEMTPEETEDIANTPSTIPPTPTVSTADILYIKAT  
HVSPSLVDALAOARALTKPSPIKSTTSAPSYVAIPHSTASPPDIATHTPO  
PPAKRSRISEKSPGASVSTFEQKKEALLRKMEIRHTELKLELESTIE  
VOEKROELHAIELRAKAMRLNGAIDESIDINORVLRONGMAVATNDERFE  
MOMPFGVLMGPFOTFLISLIVPEFAPETEGTSFVFETILLVCSIVNIIHVEF  
HRRFTLAGALYIPATILVFLTSVAVCCITGMVSTLVLIGFVDSLFRSLITVAT  
LITACALITRACITVWLFRAAPNGIKGLVOVLEGDRTSKITGAHAISTTTT  
TTTTTTTNEV"  
complement(24287..29464)  
/gene="M106.3"  
complement(join(24287..24394,24693..24749,24799..24885,  
25598..25732,25820..25908,26455..26545,27033..27442,  
27289..27439,27941..28087,28139..28563,29365..29464))  
/gene="M106.3"  
/note="Weak similarity with quinone oxidoreductase; CDNA  
EST EMBL:cl3104 comes from this gene; CDNA EST yk64f3.5  
comes from this gene"  
/codon\_start=1  
/protein\_id="CA87050.1"  
/db\_xref="PID:e1351704"  
/db\_xref="PID:g3878713"  
/db\_xref="GI:3878713"  
/db\_xref="SPTREMBL:009593"  
/translation="MRFQKQMTSSSSNAHQPLPADNKYIVYKRGENNATIDCF  
DVYFIDSPSEKIDTSDQIVRTFLSLVDPAQRKMPSTGVDLGPEIGEPVGMG  
GVVERYSACLEVDGLVGCIRLMTWIRFYVCDSDLVKVLPSRNSPSPVISC  
AGLSGITALGIRKALIDRSRPOITVYSAAAGSCGLAGQIRIGCSVYIGICSD  
DKCTVLRKRGFNDTIYKTEENVSERGIAGLAPGICIDYMDNVGVSIDVITRMNNG  
RVVLGGIANYNDLPPLPLPHTHTTKIKERNIQGRERYVLMYKKEIDAVNAQLEW

```

gene
complement(32748..37364)
/gene="M106.4"
complement(join(32748..32852,32904..33452,33507..34096,
34145..34504,34647..35159,35284..35344,36894..36998,
37199..37261,37332..37364))
/gene="M106.4"
/notes="similar to GMP synthase: cDNA EST EMBL:M88983 comes
from this gene; cDNA EST EMBL:D27937 comes from this gene;
cDNA EST EMBL:D27937 comes from this gene; cDNA EST
EMBL:D37506 comes from this gene; cDNA EST yk20016.5 comes
from this gene; cDNA EST yk208d1.5 comes from this gene;
cDNA EST yk356d1.5 comes from this gene; cDNA EST
yk486f1.5 comes from this gene"
/codon_start=1
/protein_id="CA87052.1"
/db_xref="PID:e1351706"
/db_xref="GI:3878715"
/translation="MKFSQVYPCSPKRYKIGAROSKRRHCRNLPSQLORAKTRAVSK
VSEHTKRSSMLDINEDSOHSTNKAPPKAPEDRPDSANNNGSHVTLVPEKY
SSGERIALIDGGAQGYIDRVRRLVLOSEPPLNTARTILEGKGIITISGPN
SVFEAPASIDELITCGLPVGLICGYQLMKNLGVTRHEDACETQVDTISV
HLFNGHKTETVLTHGDSVSATVAPDFKANKGSHVACINENRKLAYGFQPEP
DLTNGTKMEFLEFVYVCCGFTIONEGCISLISYIDKRLVAVSGQSAV
CAALTRALGPRVTAIHIDNGFMRESDAVERKSLADZLEIHYNGVTRSTSTH
CKDGVALDECDPEMKRKIIITFIKVDKIDINHDPEFLAOGTLPREDLESA
SALAGHADPTIKTHNDPEFLVRELKGVPEPLDFKHDEPRERQEDRESYOR
HFFPGGLATIKLCAKSDKREPLMENGPMNGVGRITIDYEQFISALNPEQIYEH
LOFEQARNVNLISERDQLAEQOTFQISALILEIKTVGVDGRSRYAVALSTEQ
RPIPKLFLAVASVLPKLFHGVNRYVAFSGIESIEDLRTYLVPHVTLQMAH
IANNILFGRVASDGTOLPNVGHKIQOMPVLLPVDPEDRMISYRSYIRLFEVT
SDMTGGAIPGVHLPBELIEMDKAIRNVGISRVLLDMCKRPGTTEW"
3787..38971
/gene="cap-2"
join(3787..37789,37838..37955,38005..38098,38147..38260,
38313..38463,38516..38620,38666..38791,38870..38971)
/gene="cap-2"
/notes="similar to F-actin capping protein beta subunit;
cDNA EST EMBL:T00119 comes from this gene; cDNA EST
EMBL:T00118 comes from this gene; cDNA EST yk323a2.5 comes
from this gene; cDNA EST yk456a6.5 comes from this gene;
cDNA EST yk471e9.5 comes from this gene"
/codon_start=1
/protein_id="CA87051.1"
/db_xref="PID:e1351705"
/db_xref="GI:3878714"
/db_xref="GI:3878714"
/db_xref="SWISS-PROT:P34686"
/translation="MGEOQLDCLALDMLRPLPHOCDKNLTLDLIDCPHLVDLLSTID
OPKRAADRECKQYLCDYNRDGDSYSPMNTYDPLPDLQDLSERKRRKETANA
AFESTRDLFBGVSATFWDLDNGFAGVITLKEGDARNTICMDSIHAEITIER
AQAHYKLTSTIMLQTNKSSGVNMGSLTRQHEMDAPINDONTNANMANRGN"
OESKMRILTINEIYFGKTKKMSDLSTEROSLEKODEIVRELNNANRGN"
join(complement(236949.1:1212..1290),
complement(236949.1:1036..1156),
complement(236949.1:203..989),
complement(236949.1:102..146),complement(39640..39973),
complement(39385..39590))
/gene="cyp-4"
/notes="similar to cyclophilin; cDNA EST yk464d7.5 comes
from this gene"
/codon_start=1
/protein_id="CA87053.1"
/db_xref="PID:e1351707"
/db_xref="GI:3878716"
/db_xref="GI:3878716"
/translation="MGKROHOKDKLYLTSEKSIIGGHKDDGTBLQRAQFRLPINH
CSLSLPREDPYCARSGEFDLTATVPYIKRKGKPCNPYAADLILHAKDKSGED
KFRCPVTRFTDHSIIAIAISGNTSHEAVQELNKNHKLDTLVDVPSFRAIDID

```

```

LODPNHLKFNMEQFELHYKLDLKTSEELKKEKDAKDKPFYIRRMNACKSVLDDLDK
EIVPKKSTVEDETDELDEINAAHYSOGKAAFTSTVAPVTSNKAANDNTRVRSY
KMAFVRLVTNFGPLNLELFAKPKACENITHCNSNYNNTKHLIKNMQLGGD
PLTGHGGEISIMDKFSDSELTSGSHDARGVLSAMKNGSNTNGSFEFTFPCYKLDRA
KTIIGRLVGGODLTITTEKLETEEGTGVPMVSVYIMAEVFDVEEAEKQVQERA
ELIKTSKDAISLANKKAKETATPEAAGTGVGKMSAAAVNKRQKMEVPLEAAK
KTRARACLGPFKRW"
join(246794.1:39146..39239,246794.1:39326..39409,
246794.1:39569..40500,246794.1:40552..40703,
246794.1:41164..41269,246794.1:41319..41452,
AL031266.1:107..112,AL031266.1:912..1068,453..614,
5546..5638,5736..6328,6853..7102,8138..8410,8672..8797,
9022..9321,9504..9776)
/gene="M106.1"
/notes="similar to chromosome segregation protein: cDNA EST

Query Match
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2826 AATTTCTAAATGATTTT 2846
|||||
Db 27521 AATTTCTAAATGATTTT 27501

RESULT 12
AC005041
LOCUS AC005041 191356 bp DNA PRI 24-MAR-1999
DEFINITION Homo sapiens clone NH0523H20, complete sequence.
ACCESSION AC005041
VERSION 94508118
NID AC005041.2 GI:4508118
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 191356)
AUTHORS Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 191356)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 191356)
REFERENCE Waterston,R.H.
AUTHORS Direct Submission
Submitted (24-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Mar 24, 1999 this sequence version replaced gi:3309088.
FEATURES
source
1..191356
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0523H20"
BASE COUNT 46280 a 47882 c 48007 g 49187 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 21; DB 42; Length 191356;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1597 CAGCCTCCCAAACTCTTC 1617
|||||
Db 43335 CAGCCTCCCAAACTCTTC 43355

RESULT 13

```

[illegible]

gene  
4412.5473  
/gene="cara"  
/EC\_number="6.3.5.5"  
/function="arginine biosynthesis"  
/note="alternate gene name: cpa"  
/codon\_start=1  
/transl\_table=11  
/product="Carbamoyl-phosphate transferase-arginine  
(subunit A)"  
/protein\_id="CAB12980.1"  
/db\_xref="PID:e1183143"  
/db\_xref="PID:g2633477"  
/db\_xref="GI:2633477"  
/db\_xref="SWISS-PROT:P36838"  
/translation="MEGLVLEDTGTSFSGELDHGHCNGTGEAVFTGMTGYOVLDPDS  
YKGOIVFTYPLIGNYGINKEDFSKPOYKAYAYVEACDHSHEAYSLKEYLOK  
NIPILTHDTRAVYKIRANGTMTATVASKGATIALQPEVNAQASQESTGDC  
NKHILIDFGIKKSLASSLVKRGCVTVVPTQOMAVYNIKPDGIVLNDKDPFAIQ  
PYLGRKISKIIISNFTLIGICLHQLIALFGGTFELPGHGANHPVIDRKRYFT  
SONHSYVDEOSINEELTIRFHHVNDTSVEGLAKKLPMVSQVCFPEAPAAESEW  
IEDYLNKVVIPAREIAHA"  
5466.8549  
/gene="carb"  
5466.8549  
/gene="carb"  
/EC\_number="6.3.5.5"  
/function="arginine biosynthesis"  
/note="alternate gene name: cpa"  
/codon\_start=1  
/transl\_table=11  
/product="Carbamoyl-phosphate transferase-arginine  
(subunit B)"  
/protein\_id="CAB12981.1"  
/db\_xref="PID:e1183144"  
/db\_xref="PID:g2633478"  
/db\_xref="GI:2633478"  
/db\_xref="SWISS-PROT:P18185"  
/translation="MPKDTSSISLVIGSGPIIIQGAAEFDYSGTGCIALKEEGRV  
ILVNSPATIMTDEAFDEIYEPFLAEISALAIKKERPDGLANLGGOTLAAVEL  
ETGVYKHEGKILGTSVETIOKGDREKFRSLAMELKOPPESTIVNEDALHAE  
SIGFVILIRPATILGKGGIAPDEKATITAKOLASPLNQCLVENSIGKEIEY  
EVMRDSNNTCIVCMENIDVGVHTGSIYVAPQTLTDEYOMLRASLTIIISALD  
VGGGNIOTFALDPESKOYVLEVNVRSSSLASKAGYPLAKMAALAVGYLDEL  
KNPLTGSYASFEPLADYVYKFPMPKDFKNADKRGTKKATGEVMAIERNEEA  
IOKAGASLEKINCHLPELISGLSTDTLMDLAIPDDRPVYVMEILSRVSIIDHIE  
KTKIDPFLHFDNITIKLENLMKMGSLSELLKAKKEGSDATISLISKTEEV  
RLRKEMGITPSFKIVTCAAEFDKATINFTSTFGEIDGDISREKRRALITSGPI  
RIQGVVEVDSYAVHGLTLOEAGFTIMINNPEVSDYEIADRLYEPMPTTEIVN  
VREQENIDFAIVQGGTAIINAALERAAGITLGTSTQTLVDLDRQFOLDDELG  
LKHKAGEIATYKEEASAKSEIGYVPLIRPSYVIGMGMIIVDSQASOLINDSDM  
PYLIDQYVGKKEVIDLISDGEVFIPTYTEHERAGVHSGDSFALIPGSTISGL  
OQGDMAOKIARAKISFKGINIQCVIDNGNILEVNPRASTRVPSKMGVPRAT

Query Match 0.7%; Score 20; DB 1; Length 216750;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1858 GCAATATTCGCCAGGAGC 1877  
|||||  
Db 37659 GCAATATTCGCCAGGAGC 37640

RESULT 14  
BTU36445 2984 bp mRNA MAM 09-FEB-1996  
LOCUS Bos taurus calcium-activated chloride channel mRNA, complete cds.  
DEFINITION U36445  
ACCESSION 91184065  
NID U36445.1 GI:1184065  
KEYWORDS

SOURCE  
ORGANISM Bos taurus  
REFERENCE  
AUTHORS 1 (bases 1 to 2984)  
Cunningham, S.A., Awayda, M.S., Budlen, J.K., Ismailov, I.I.,  
Arrate, M.P., Berdier, B.K., Benos, D.J. and Fuller, C.M.  
TITLE Cloning of an epithelial chloride channel from bovine trachea  
JOURNAL J. Biol. Chem. 270 (52), 31016-31026 (1995)  
MEDLINE 96125078  
REFERENCE 2 (bases 1 to 2984)  
AUTHORS Benos, D.J.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,  
University of Alabama at Birmingham, BHSB 706, Birmingham, AL  
35294, USA

FEATURES  
source location/Qualifiers  
1.2984  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19.2730  
/note="epithelial chloride channel"  
/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="PID:g1184066"  
/db\_xref="GI:1184066"  
/translation="MWPRLYTILFLTLHLPLGKSSWNLINNGYDGVIVAINSPVE  
DEKLONIKEMVTEASTLYLHATKRRYFRNVSILIPMTWKSSTLPRKQESTDOAE  
VIVANPYLKHGDDPYTLQYGRGCGKQYIHTEPFLTNLNPITGSRGARVHEWAL  
RMGJFDEYNGDDEPYISRNRTIEATRCSTHITGVNVCOCGSGCITPRCARDSQTL  
YEACFTPEKSGTARRESIMOSLSHSTEECTKEHTEHAPNOLNCKNCKSTWDV  
MSNDPONTSPMTMNPPTOPTGLSLSKOVNCTLVLDKSGMSSEDELFPNNOAEL  
FLIOIEKSGVGNWTPDSVAIEKNNLTKITDDVYENITANLQENAGGTSICGLK  
AGFOALIOSQOSTSGSEITLITLDEBNDHSCIEVKSGVYIHTALGPSAAKELET  
LSDMTGGRFYANKDINGLITNAFHSRISGSIQOITQLSKALAITKKMVGATVP  
VDSIIGNDFEVVMTIKKPELILADDPGKKYKTSDEKEDLNIHSARLRIPGIAETG  
TWYSLNNHNASPOLITVYVTRARSPTTPVTAAHNSOMTAHPSPIVYAOVSOG  
FLPVLGINVTALITTEGSHOYTLILMNGACADYVKNNGISYRFTDYRGGRSLKY  
HAERANNTARSLKQPOKKAITYGIEYKNGKIIINPRPEYKDLARAEIDPFSRIS  
GGSTVSGAPGNHPSVLPKNTIIDLEKKEEDHIOQSWTAPANVLADKANSYITRI  
SKSLDIDKQDNDMTLVNTSSLKREKAGSDENFEKPEPFIENGTEYIAVQAINEA  
NLTEVSLSIAQAKIFPIPEDSVALGKISAINLAIKALMIISIV"

BASE COUNT 1027 a 583 c 561 g 813 t  
ORIGIN

Query Match 0.7%; Score 20; DB 3; Length 2984;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 481 GCATTTCGATGAGTGCGC 500  
|||||  
Db 472 GCATTTCGATGAGTGCGC 491

RESULT 15  
AB025615/c 83544 bp DNA PLN 07-APR-1999  
LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone: K5K13,  
DEFINITION complete sequence.  
ACCESSION AB025615  
NID 94589421  
VERSION AB025615.1 GI:4589421  
KEYWORDS HTG.  
SOURCE Arabidopsis thaliana (strain: Columbia) DNA, clone: lib: Mitsu1 P1  
clone: K5K13.  
ORGANISM Arabidopsis thaliana; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core

eudicots: Rosidae; eurosids II; Brassicales; Brassicaceae;  
Arabidopsis.

REFERENCE 1 (sites)  
AUTHORS Nakamura, Y.  
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. II  
JOURNAL Unpublished (1999)  
REFERENCE 2 (bases 1 to 83544)  
AUTHORS Nakamura, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-APR-1999) to the DDBJ/EMBL/GenBank databases.  
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of  
Gene Structure 2, 1532-3, Yana, Kisarazu, Chiba 292, Japan  
(E-mail: ynakamura@kazusa.or.jp, Tel: +81-438-52-3935,  
Fax: +81-438-52-3934)

FEATURES  
Source Location/Qualifiers  
1..83544  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="3"  
/clone="K5K13"  
/clone\_lib="Mitsui P1"

BASE COUNT 28677 a 14395 c 13562 g 26910 t  
ORIGIN

Query Match 0.7%; Score 20; DB 7; Length 83544;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2828 ATTTCTAAATGTAATTTA 2847  
|||||  
Db 82992 ATTTCTAAATGTAATTTA 82973

Search completed: August 6, 1999, 12:09:32  
Job time: 8258 sec





```

PN W09639419-A1.
PD 12-DEC-1996.
PF 06-JUN-1995; U07289.
PR 06-JUN-1995; WO-U07289.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Yu G;
PFI WPI: 97-043054/04.
DR P-PSDB: W06548.
PT Human colon specific genes and their expression products - detection
PT of which, in non-colon tissue samples, can be used as indication of
PT colon cancer metastasis
PS Claim 1: Fig 5: 60pp: English.
CC 13 cDNA clones (T45880-92), most of them partial clones, correspond
CC to human colon specific genes, designated CSG1, CSG2, etc., that
CC are primarily expressed in tissues derived from the colon. CSG1
CC and CSG10 show reduced expression in colon cancer cells as compared
CC to that in normal cells; the remaining genes are overexpressed in
CC colon cancer. The partial cDNA sequences can be used to isolate
CC full-length clones and genomic clones including the complete gene.
CC CSG nucleic acids can be used to produce CSG polypeptides (see also
CC W06543-53) in transformed host cells, as probes to detect disorders
CC of the colon, partic. colon cancer and colon cancer metastasis, and
CC in gene therapy.
SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 31.1%; Score 33; DB 1; Length 878;
Best Local Similarity 100.0%; Pred. No. 6.9e-08;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCATTACATTTAAATATGTGGAAGTGG 33
DB 717 GGCATTACATTTAAATATGTGGAAGTGG 749

RESULT 3
ID V16672 standard; cDNA; 878 BP.
AC V16672;
DT 22-JUN-1998 (first entry)
DE Polynucleotide sequence of a colon-specific gene.
KW Colon-specific gene; probe; detection; expression; human;
KM diagnostic assay; colon cancer; antibody; screening; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 2..685
FT /tag= a
FT /note= "no stop codon given"
PN US5733748-A.
PD 31-MAR-1998.
PF 06-JUN-1995; 469667.
PR 06-JUN-1995; US-469667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen C, Yu G;
PFI WPI: 98-229823/20.
DR P-PSDB: W46879.
PT Colon-specific nucleic acids - useful as probes for detecting colon
PT cancer metastases
PS Claim 15; Fig 5b-B; 51pp; English.
CC V16668-81 represent polynucleotide sequences of partial or full-length
CC cDNA clones of colon-specific genes. The polynucleotides can be used
CC as probes to detect expression of the corresponding human genes, e.g. in
CC diagnostic assays for detecting micrometastases of colon cancer.
CC Recombinant cells containing the polynucleotides can be used to
CC produce the polypeptides, in order that antibodies can be raised and
CC used in further screening or diagnostics.
SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 31.1%; Score 33; DB 1; Length 878;
Best Local Similarity 100.0%; Pred. No. 6.9e-08;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 GGCATTACATTTAAATATGTGGAAGTGG 33
DB 717 GGCATTACATTTAAATATGTGGAAGTGG 749

RESULT 4
ID X13099 standard; DNA; 14286 BP.
AC X13099;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:162.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN W09850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046635.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
PFI WPI: 99-045171/04.
DR New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 907-914; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 14286 BP; 4262 A; 2884 C; 2303 G; 4821 T;

Query Match 17.0%; Score 18; DB 1; Length 14286;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 77 TCAGATAAATTAATAATAA 94
DB 10683 TCAGATAAATTAATAATAA 10666

RESULT 5
ID N90490 standard; DNA; 18177 BP.
AC N90490;
DT 30-OCT-1989 (first entry)
DE DNA of human retinoblastoma gene exons with flanking regions.
KW Human; retinoblastoma gene; flanking regions; cDNA;
KW screening; tumours; probes; exons.
OS Homo sapiens (human).
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
FT exon 1 533..669
FT intron (2.9 kb) 1008/1009
FT exon 2 1318..1444
FT intron (33 kb) 1712/1713
FT exon 3 1995..2110
FT intron (1.6 kb) 2363/2364
FT exon 4 2559..2678
FT intron (2.1 kb) 2933/2934

```

```
FT      exon 5      3227..3265
FT      Intron (1.4 kb) 3549/3550
FT      exon 6      3695..3762
FT      Intron (11.5 kb) 3983/3984
FT      exon 7      4285..4395
FT      Intron (1.8 kb) 4638/4640
FT      exon 8      4909..5051
FT      Intron (1.8 kb) 5248/5249
FT      exon 9      5480..5561
FT      Intron (1.9 kb) 5734/5735
FT      exon 10     6000..6109
FT      Intron (1.0 kb) 6381/6382
FT      exon 11     6690..6768
FT      Intron (3.2 kb) 7061/7062
FT      exon 12     7375..7462
FT      Intron (3.1 kb) 7765/7766
FT      exon 13     7996..8112
FT      Intron (1.8 kb) 8349/8350
FT      exon 14     8680..8736
FT      exon 15     9140..9171
FT      exon 16     9252..9328
FT      Intron (1.0 kb) 9629/9630
FT      exon 17     9879..10075
FT      Intron (70.0 kb) 10322/10323
FT      exon 18     10564..10682
FT      Intron (3.0 kb) 10911/10912
FT      exon 19     11225..11370
FT      Intron (2.8 kb) 11561/11562
FT      exon 20     11809..11955
FT      Intron (4.0 kb) 12278/12279
FT      exon 21     12538..12642
FT      Intron (0.2 kb) 12948/12949
FT      exon 22     13204..13317
FT      Intron (7.4 kb) 13411..13574
FT      exon 23     13883/13884
FT      Intron (2.8 kb) 14209..14239
FT      exon 24     14641/14642
FT      exon 25     14834..14977
FT      exon 26     15488..15538
FT      Intron (1.7 kb) 15952/15953
FT      exon 27     16143..16213
FT      Polya signal 18010

W08906703-A.
PD 27-JUL-1989.
PF 23-JAN-1989: U00293.
PR 21-JAN-1988: US-146525.
PA (DRYJ) Dryja T P.
PI Dryja T P; Friend S; Yandell D W.
DR P-PSDB; P90599.
PT New human retino-blastoma gene and polypeptide(s)
PT - used for screening individuals for defective retino-blastoma
PT gene and treating these patients.
PS Disclosure: fig 6; 71 pp; English.
CC DNA encoding human retinoblastoma (RB) gene exons and flanking regions.
CC The CDNA (see P90599 and N90490) can be used to screen individuals
CC for the presence of the mutated RB gene. The RB polypeptide
CC can prevent retinoblastoma formation, and corresp. antibodies
CC can be used in tumour immunodiagnosis. Refer to patent for more
CC sequence analysis details.
SQ Sequence 18177 BP; 5765 A; 2997 C; 3172 G; 6202 T; 41 Others.
```

```
Query Match      16.0%; Score 17; DB 1; Length 18177;
Best Local Similarity 100.0%; Pred.No. 6.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 7 CACATTTTAAATAT 23
|||||
Db 6298 CACATTTTAAATAT 6282
```

RESULT 6

```
V21209_02
Continuation (3 of 17) of V21209 from base 200001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 200001 310000
WP V21209_03 300001 410000
WP V21209_04 400001 510000
WP V21209_05 500001 610000
WP V21209_06 600001 710000
WP V21209_07 700001 810000
WP V21209_08 800001 910000
WP V21209_09 900001 1010000
WP V21209_10 1000001 1110000
WP V21209_11 1100001 1210000
WP V21209_12 1200001 1310000
WP V21209_13 1300001 1410000
WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976
```

```
Query Match      16.0%; Score 17; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred.No. 6.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 78 CAGATAATATAATATA 94
|||||
Db 69638 CAGATAATATAATATA 69654
```

```
RESULT 7
Continuation (15 of 17) of V21209 from base 1400001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 200001 310000
WP V21209_03 300001 410000
WP V21209_04 400001 510000
WP V21209_05 500001 610000
WP V21209_06 600001 710000
WP V21209_07 700001 810000
WP V21209_08 800001 910000
WP V21209_09 900001 1010000
WP V21209_10 1000001 1110000
WP V21209_11 1100001 1210000
WP V21209_12 1200001 1310000
WP V21209_13 1300001 1410000
WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976
```

```
Query Match      16.0%; Score 17; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred.No. 6.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 81 ATTAATATAATATAATCA 97
|||||
Db 96454 ATTAATATAATATAATCA 96470
```

```
RESULT 8
X04502/c
ID X04502 standard; DNA; 18303 BP.
AC X04502;
DT 05-JUL-1999 (first entry)
DE Human retinoblastoma gene.
KW Retinoblastoma; RB gene; human; predisposition; diagnosis; therapy;
ds.
```

```

OS Homo sapiens.
FH Key Location/Qualifiers
FH TATA_signal 122..125
FT /tag= a
FT 533..16215
FT /tag= b
FT /note= "contains 26 introns"
FT /tag= c
FT /number= 1
FT /note= "deposited as ATCC 97927 and ATCC 97928,
FT specifically claimed in Claim 20"
FT 671..1317
FT /tag= d
FT /number= 1
FT /note= "the intron is 2.9 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 1008 and 1009"
FT 1318..1444
FT /tag= e
FT /number= 2
FT /note= "deposited as ATCC 97929, specifically
FT claimed in Claim 21"
FT 1445..1994
FT /tag= f
FT /number= 2
FT /note= "the intron is 3.3 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 1712 and 1713"
FT 1995..2110
FT /tag= g
FT /number= 3
FT /note= "deposited as ATCC 97932, specifically
FT claimed in Claim 22"
FT 2111..2558
FT /tag= h
FT /number= 3
FT /note= "the intron is 1.6 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 2363 and 2364"
FT 2559..2678
FT /tag= i
FT /number= 4
FT /note= "deposited as ATCC 97930, specifically
FT claimed in Claim 23"
FT 2679..3226
FT /tag= j
FT /number= 4
FT /note= "the intron is 2.1 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 2933 and 2934"
FT 3227..3266
FT /tag= k
FT /number= 5
FT /note= "deposited as ATCC 97934 and ATCC 97935,
FT specifically claimed in Claim 24"
FT 3267..3694
FT /tag= l
FT /number= 5
FT /note= "the intron is 1.4 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 3550 and 3551"
FT 3695..3762
FT /tag= m
FT /number= 6
FT /note= "deposited as ATCC 97933, specifically
FT claimed in Claim 25"
FT 3763..4284
FT /tag= n
FT /number= 6
FT /note= "the intron is 11.5 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 3983 and 3984"

FT exon 4285..4396
FT /tag= o
FT /number= 7
FT /note= "deposited as ATCC 97946, specifically
FT claimed in Claim 26"
FT 4397..4908
FT /tag= p
FT /number= 7
FT /note= "the intron is 1.8 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 4639 and 4640"
FT 4909..5051
FT /tag= q
FT /number= 8
FT /note= "deposited as ATCC 97945, specifically
FT claimed in Claim 27"
FT 5052..5483
FT /tag= r
FT /number= 8
FT /note= "the intron is 1.8 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 5248 and 5249"
FT 5484..5561
FT /tag= s
FT /number= 9
FT /note= "deposited as ATCC 97936, specifically
FT claimed in Claim 28"
FT 5562..5999
FT /tag= t
FT /number= 9
FT /note= "the intron is 1.9 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 5734 and 5735"
FT 6000..6109
FT /tag= u
FT /number= 10
FT /note= "deposited as ATCC 97937, specifically
FT claimed in Claim 29"
FT 6110..6690
FT /tag= v
FT /number= 10
FT /note= "the intron is 1.0 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 6381 and 6382"
FT 6691..6768
FT /tag= w
FT /number= 11
FT /note= "deposited as ATCC 97937, specifically
FT claimed in Claim 30"
FT 6769..7374
FT /tag= x
FT /number= 11
FT /note= "the intron is 3.2 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 7061 and 7062"
FT 7375..7462
FT /tag= y
FT /number= 12
FT /note= "deposited as ATCC 97938, specifically
FT claimed in Claim 31"
FT 7463..7995
FT /tag= z
FT /number= 12
FT /note= "the intron is 3.1 kb, there is a break in
FT the sequence given in the specification,
FT between nucleotides 7765 and 7766"
FT 7996..8112
FT /tag= aa
FT /number= 13
FT /note= "deposited as ATCC 97950, specifically
FT claimed in Claim 32"
FT 8113..8679
FT /tag= ab

```

```

FT      /number- 13
FT      /note= "the intron is 1.8 kb, there is a break in
FT      the sequence given in the specification,
FT      between nucleotides 8349 and 8350"
FT      exon
FT      8680..8736
FT      /tag= ac
FT      /number= 14
FT      /note= "deposited as ATCC 97950, specifically
FT      claimed in Claim 33"
FT      intron
FT      8737..9139
FT      /tag= ad
FT      /number= 14
FT      9140..9174
FT      /tag= ae
FT      /number= 15
FT      /note= "deposited as ATCC 97950, specifically
FT      claimed in Claim 34"
FT      intron
FT      9175..9251
FT      /tag= ae
FT      /number= 15
FT      9252..9328
FT      /tag= af
FT      /number= 16
FT      /note= "deposited as ATCC 97950, specifically
FT      claimed in Claim 35"
FT      intron
FT      9329..9878
FT      /tag= ag
FT      /number= 16
FT      /note= "the intron is 1.0 kb, there is a break in
FT      the sequence given in the specification,
FT      between nucleotides 9629 and 9630"
FT      exon
FT      9879..10076
FT      /tag= ah
FT      /number= 17
FT      /note= "deposited as ATCC 97950, specifically
FT      claimed in Claim 36"
FT      intron
FT      10077..10563
FT      /tag= ai
FT      /number= 17
FT      /note= "the intron is 70.0 kb, there is a break in
FT      the sequence given in the specification,
FT      between nucleotides 10322 and 10323"
FT      exon
FT      10564..10682
FT      /tag= aj
FT      /number= 18
FT      /note= "deposited as ATCC 97949, specifically
FT      claimed in Claim 37"
FT      intron
FT      10683..11224

Query Match
Best Local Similarity 16.0%; Score 17; DB 1; Length 18303;
Pred. No. 6.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      7 CACATTAAATAATAT 23
      |||||
Db      6362 CACATTAAATAATAT 6346

RESULT 9
N91687/c
ID      N91687 standard; cDNA; 1153 BP.
AC      N91687;
DT      12-MAY-1990 (first entry)
DE      Human pituitary cDNA encoding 7B2 preprotein
KW      7B2 preprotein; human pituitary; small cell lung tumour.
OS      Homo sapiens.
FH      Key
FH      Location/Qualifiers
FT      cds
FT      28..663
FT      /tag= a
FT      925..925
FT      /tag= b
FT      1128..1128
FT      /tag= c

```

```

PN      EP-315254-A.
PD      10-MAY-1989.
PF      28-OCT-1988; 202395.
PR      30-OCT-1987; NL-002590.
PA      (UYKA-) Sticht Kathol Univ.
PI      Martens GJM;
DR      WPI; 89-139252/19.
DR      P-PSDB; P91077.
PT      Recombinant human 7B2 protein
PT      - used as marker for small cell lung- and endocrine
PT      tumours and for producing antibodies for immuno-analysis
PS      Figure 1; page 12; 15pp; English.
CC      It encodes human 7B2 preprotein with an 18-26 amino acid signal peptide
CC      sequence and a 185 amino acid mature 7B2 protein. 7B2 is a marker for
CC      classic small cell lung tumours which may be detd. using labelled RNA or
CC      DNA probes or the Abs.
SQ      Sequence 1153 BP; 323 A; 241 C; 283 G; 306 T;

Query Match
Best Local Similarity 15.1%; Score 16; DB 1; Length 1153;
Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      4 ATTCACATTTAAAA 19
      |||||
Db      887 ATTCACATTTAAAA 872

RESULT 10
Q33628
ID      Q33628 standard; DNA; 174 BP.
AC      Q33628;
DE      02-FEB-1993 (first entry)
DE      Downstream sequence of microsatellite from clone GBKAS.
KW      PCR; selection; primers; OPTIPRM; breeding; cattle; parentage;
KW      genetic mapping; traits; amplification; ss.
OS      Bos taurus.
PN      WO9213102-A.
PD      06-AUG-1992.
PF      15-JAN-1992; U00340.
PR      15-JAN-1991; US-642342.
PA      (GENM-) GENMARK.
PI      Georges M. Massey JM;
DR      WPI; 92-284684/34.
PT      Polymorphic bovine DNA markers - used in genetic identification,
PT      gene mapping, and selective breeding
PS      Table 7; page 180; 517pp; English.
CC      The sequence is that downstream of a bovine microsatellite sequence
CC      obtd. by screening a library of bovine MboI DNA fragments of between
CC      250 and 500 bp with an (AC)15 and a (TC)15 oligonucleotide probe.
CC      One out of 50 clones cross-hybridised. Assuming independent
CC      distribution of microsatellites and MboI sites, the frequency of
CC      (T6)n >9 microsatellites in the bovine genome is estimated at >100.
CC      000. The sequence information for ca. 230 such bovine microsatellites
CC      is summarised in the specification and indexed herein (see below).
CC      The sequences upstream and downstream of the microsatellite sequence
CC      were used to generate the required PCR primers for in vitro
CC      amplification of the corresp. microsatellite (using the program
CC      OPTIPRM). The microsatellites may be used to identify individuals,
CC      for parentage testing, and in the genetic mapping of economic traits,
CC      loci, or genes involved in the determination of economically important
CC      traits esp. in cattle, to allow selective breeding.
CC      See also Q33501-34437.
SQ      Sequence 174 BP; 77 A; 31 C; 19 G; 47 T;

Query Match
Best Local Similarity 15.1%; Score 16; DB 1; Length 174;
Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      79 AGATAAATRAAATAAA 94
      |||||
Db      129 AGATAAATRAAATAAA 144

```

```

RESULT 11
Q57356/c
ID 077356 standard; DNA: 194 BP.
AC 077356;
DT 22-NOV-1994 (first entry)
DE Human genome fragment (Preferred).
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
OS detection; homology; human; adrenal tissue; ds.
PN WO9401548-A.
PD 20-JAN-1994.
PF 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014857.
PA (MED1-) MEDICAL RES CONCL.
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI Sidsen DR, Starkey M;
PT WPI: 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PS for genetic analysis and mapping.
PS Claim 1; Page 484; 616pp; English.
CC Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (076401-077613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
SQ Sequence 194 BP; 66 A; 22 C; 20 G; 86 T;

Query Match 15.1%; Score 16; DB 1; Length 194;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ATTGCACATTTTAAAAA 19
DB 142 ATTGCACATTTTAAAAA 127

RESULT 12
T01862/c
ID T01862 standard; CDNA: 1153 BP.
AC T01862;
DT 05-MAR-1996 (first entry)
DE Human 7B2 pre-protein CDNA.
KW Pre-protein 7B2; protease-binding portion; inhibition; AIDS;
KW acquired immune deficiency syndrome; hormonal disorders;
KW blood clotting cascade disorders; protease cleavage site; variant;
KW pro-hormone convertase; PC2; paired basic amino acids; ss.
OS Homo sapiens.
PH Key Location/Qualifiers
FT signal_peptide 29..106
FT /*tag- a
FT /note- "encodes pre-region"
FT mat_peptide 107..661
FT /*tag- b
FT /product- 7B2-protein
FT poly_a_signal 926..931
FT /*tag- c
FT poly_a_signal 1128..1133
FT /*tag- d
FT poly_a_site 1153
FT /*tag- e
FT /note- "start of polyA tail of undefined length"

PN N9400032-A.
PD 01-AUG-1995.
PF 07-JAN-1994; 000032.
PR (UYKA-) UNIT STITCHING KATHOLIEKE.
PA Martens GJM;
PI WPI: 95-268117/35.

```

```

DR P-PSDB: R83555.
PT Use of new and known 7B2 proteins - as protease inhibitors, esp. for
PT treating AIDS, hormonal disorders and blood clotting disorders
PS Claim 24; Fig 1; 35pp; Dutch.
CC Protein 7B2, its protease-binding fragments and derivs in which the
CC protease cleavage site is altered are all useful for inhibiting the
CC activity of pro-hormone convertase enzymes (esp. PC2). Release of
CC hormones such as insulin, opiomelanocortin, enkephalins or
CC endorphins from their pro-hormones can be inhibited for treating
CC hormonal disorders. Inhibition of furin-induced release of blood
CC clotting factors from pro-proteins is useful for treating disorders
CC of the blood clotting cascade and inhibition of furin-induced
CC cleavage of gp160 is useful for treating AIDS. Sequences coding for
CC protein 7B2 or its derivs. can be used for gene therapy of the
CC above diseases.
SQ Sequence 1153 BP; 323 A; 241 C; 282 G; 307 T;

Query Match 15.1%; Score 16; DB 1; Length 1153;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ATTGCACATTTTAAAAA 19
DB 888 ATTGCACATTTTAAAAA 873

RESULT 13
V57903
ID V57903 standard; DNA: 237326 BP.
AC V57903;
DT 21-DEC-1998 (first entry)
DE Hereditary haemochromatosis subregion from an HH affected individual.
KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NP13; NP14; Rorret; BTF1; BTF3;
KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene; ss.
OS Homo sapiens.
PN WO9814466-A1.
PD 09-APR-1998.
PF 30-SEP-1997; U17658.
PR 07-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGENITOR INC.
PI Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK;
DR WPI: 98-240014/21.
PT Hereditary haemochromatosis gene products - used to develop products
PT for the diagnosis and treatment of hereditary disorders in iron
PT metabolism
PS Claim 1; Fig 9; 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. The present sequence
CC represents a hereditary haemochromatosis subregion from an hereditary
CC haemochromatosis (HH) affected individual. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BTF genes, which are homologues of the milk
CC protein butyrophilin (BT), and can be used in the production of agonists
CC and antagonists of BTF function. Also described are: (1) a Rorret gene
CC which can be used to develop products for the study, diagnosis and
CC treatment of lupus and Sjogren's syndrome; and (2) NP13 and NP14 genes
CC which are homologues of a type 1 sodium transport gene, and can
CC similarly be used for hypophosphatemia.
SQ Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T;

```

Query Match 15.1%; Score 16; DB 1; Length 237326;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CATTAAAAATTATG 24

DB 109190 CATTAAAAATTATG 109205

RESULT 14

ID V75590 standard; DNA; 418 BP.

AC V75590;

DE 16-MAR-1999 (first entry)

STAPHYLOCOCCUS aureus contig SEQ ID #1279.

Computer readable medium; vaccine; S. aureus infection; immunodetection;

cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

skin infection; surgical wound infection; scalded skin syndrome;

KM toxic shock syndrome; ds.

OS Staphylococcus aureus.

FT Key Location/Qualifiers

FT misc\_feature 301..360

FT /tag= a

FT /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

EP-786519-A2.

PD 30-JUL-1997.

PF 07-JAN-1997; 100117.

PA 05-JAN-1996; US-009861.

PR (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,

PI Rosen CA;

PI WPI: 97-374922/35.

DR Polynucleotide(s) and proteins derived from Staphylococcus aureus

PT stored on computer readable medium and used in the production of

PS anti-S. aureus vaccines

PS Claim 1; Page 1903-1904; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences

of the invention. The DNA sequences are recorded on a computer readable

medium, preferably selected from a floppy or hard disk, random access

memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

the S. aureus DNA sequences allows putative functions to be assigned so

that protein-encoding or regulatory regions of commercial, therapeutic or

industrial importance can be obtained. Specifically, sequences which are

likely to encode antigens have been identified and these polypeptides can

be used in a vaccine composition against S. aureus infection. The

polypeptides can also be used in a kit for the immunodetection of

S. aureus in a sample. S. aureus is implicated in numerous human diseases,

including cellulitis, eyelid infections, food poisoning, osteomyelitis,

skin and surgical wound infections, scalded skin syndrome, toxic shock

syndrome, etc. Organisms transformed with the DNA sequences can be used

for recombinant production of the polypeptides. The new DNA sequences

(and their fragments) are useful as primers or probes for isolating

homologues of any of the S. aureus DNA sequences contained on the

computer readable medium.

CC Sequence 418 BP; 151 A; 40 C; 51 G; 115 T;

Query Match 15.1%; Score 16; DB 1; Length 418;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 AATATAAATTAATCAT 98

DB 174 AATATAAATTAATCAT 189

RESULT 15

ID X13065/c

AC X13065; standard; DNA; 32768 BP.

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:128.

KM Enterococcus faecalis; contig; detection; Enterococcal infection;

KM vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN WO9850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 008985.

PR 14-NOV-1997; US-066009.

PR 06-MAY-1997; US-044031.

PR 16-MAY-1997; US-046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

PI WPI: 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

used to develop products for the detection of Enterococcus and for

use in vaccines for prevention or attenuation of Enterococcus

infection.

PS Claim 1; Page 767-783; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it

982 nucleotide sequences isolated from the Enterococcus faecalis genome.

X12938 to X13919 represent these nucleotide sequences which are primary

nucleotide sequences, also known as contigs. The computer-based system

can identify fragments of the Enterococcus faecalis genome with

commercial importance. The products can be used to detect the presence

of Enterococcus faecalis in samples. They can also be used for

diagnosing Enterococcal infection in an animal and monitoring

progression of disease, and for identifying agents which can be used to

modulate the growth or pathogenicity of Enterococcus faecalis, or

another related organism, in vivo or in vitro. In particular the

polypeptides encoded by the Enterococcus faecalis nucleotide sequences

can be used in vaccines to prevent or attenuate an Enterococcal

infection.

CC Sequence 32768 BP; 9529 A; 6883 C; 5611 G; 10720 T;

Query Match 15.1%; Score 16; DB 1; Length 32768;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AATTTAAATTAATG 25

DB 13013 AATTTAAATTAATG 12998

Search completed: August 5, 1999, 17:41:48  
Job time: 6343 sec

Mon Aug 9 13:26:54 1999

us-09-049-696-17\_1.rng

---

Page 8



GenCore version 4.5  
Copyright (c) 1993 - 1998 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:41:48 ; Search time 650.92 Seconds  
(without alignments)  
1081.224 Million cell updates/sec

Title: US-09-049-696-18

Perfect score: 2813  
Sequence: 1 GAAATCAGAGGAGATGTAC.....AATAATCATTCATCCTTA 2813

Scoring table: OLIGO.NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685	24.4	878	1 T45884	Human colon specifi
2	685	24.4	878	1 T45884	Poly nucleotide seq
3	164	5.8	255	1 T22483	Human gene signatu
4	19	0.7	710	1 Q06308	Sequence of DNA fr
5	19	0.7	2184	1 Q06301	Sequence encoding
6	19	0.7	2100	1 V59671	Human secreted pro
7	19	0.7	447	1 V90187	EST clone DH318. N
8	19	0.7	299	1 V89358	EST clone C1493. N
9	18	0.6	3642	1 N50364	DNA fragment encod
10	18	0.6	3474	1 N80908	Sequence of Herpes
11	18	0.6	897	1 Q31988	Gutline p1g MBP-1 g
12	18	0.6	3472	1 Q48497	Glycoprotein B (GB
13	18	0.6	735	1 Q73229	Soluble human inte
14	18	0.6	2040	1 T43949	Sequence flanking
15	18	0.6	1386	1 T63545	Rat pheromone rece
16	18	0.6	3473	1 T93651	Herpes Simplex vir
17	18	0.6	716	1 V68834	DNA molecule encod
18	18	0.6	14286	1 X13099	Enterococcus faeca
19	17	0.6	18177	1 N90490	DNA of human retin
20	17	0.6	843	1 Q10677	Modified hepatitis
21	17	0.6	2287	1 Q13802	Non-receptor like
22	17	0.6	1560	1 Q29613	Encodes carp TLAla
23	17	0.6	2155	1 Q42418	IDUA - exons I and
24	17	0.6	4475	1 Q45995	Clone of recombin
25	17	0.6	13104	1 Q46852	Porcine sodium ion
26	17	0.6	4529	1 Q46121	Human brain Expres
27	17	0.6	314	1 Q60964	Endothelin-1 (ET-1
28	17	0.6	420	1 Q57632	Megakaryocyte pote
29	17	0.6	2129	1 Q63972	Human gene signatu
30	17	0.6	267	1 T20069	Human endothelein-1
31	17	0.6	713	1 T31013	Human gene signatu
32	17	0.6	276	1 T26704	Oestrogen-regulate
33	17	0.6	2404	1 T33220	Continuation (6 of
34	17	0.6	80073	1 T38840.5	Human alpha-L-idur
35	17	0.6	2067	1 T71754	Expression augment
36	17	0.6	14507	1 T73568	Human CAK1 antigen
37	17	0.6	2138	1 T91079	Partial sequence o
38	17	0.6	2404	1 T99071	Aspergillus oryzae
39	17	0.6	2454	1 T97392	Human endosulfine
40	17	0.6	1199	1 V45211	Methanococcus jann
41	17	0.6	110000	1 V21209_00	Continuation (3 of
42	17	0.6	110000	1 V21209_02	Continuation (4 of
43	17	0.6	110000	1 V21209_03	

RESULT 1  
T45884  
ID T45884 standard; cDNA: 878 BP.  
AC T45884;  
AT 13-MAR-1997 (first entry)  
DE Human colon specific gene CSG5 cDNA partial clone.  
KW Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;  
OS Homo sapiens.  
FH key  
FT cds  
FT Location/Qualifiers  
FT 2..691  
FT /\*tag= a

## ALIGNMENTS

Query Match 24.4%; Score 685; DB 1; Length 878;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Sequence 878 BP: 257 A; 179 C; 188 G; 241 T;  
CSG nucleic acids can be used to produce CSG polypeptides (see also  
W06545-53) in transformed host cells, as probes to detect disorders  
of the colon, partic. colon cancer and colon cancer metastasis, and  
in gene therapy.  
CC full-length clones and genomic clones including the complete gene.  
CC to human colon specific genes, designated CSG1, CSG2, etc., that  
are primarily expressed in tissues derived from the colon. CSG  
CC and CSG10 show reduced expression in colon cancer cells as compared  
CC to that in normal cells; the remaining genes are overexpressed in  
CC colon cancer. The partial cDNA sequences can be used to isolate  
CC full-length clones and genomic clones including the complete gene.  
CC CSG nucleic acids can be used to produce CSG polypeptides (see also  
W06545-53) in transformed host cells, as probes to detect disorders  
of the colon, partic. colon cancer and colon cancer metastasis, and  
in gene therapy.  
SQ Sequence 878 BP: 257 A; 179 C; 188 G; 241 T;

Query Match 24.4%; Score 685; DB 1; Length 878;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Sequence 878 BP: 257 A; 179 C; 188 G; 241 T;  
CSG nucleic acids can be used to produce CSG polypeptides (see also  
W06545-53) in transformed host cells, as probes to detect disorders  
of the colon, partic. colon cancer and colon cancer metastasis, and  
in gene therapy.  
CC full-length clones and genomic clones including the complete gene.  
CC to human colon specific genes, designated CSG1, CSG2, etc., that  
are primarily expressed in tissues derived from the colon. CSG  
CC and CSG10 show reduced expression in colon cancer cells as compared  
CC to that in normal cells; the remaining genes are overexpressed in  
CC colon cancer. The partial cDNA sequences can be used to isolate  
CC full-length clones and genomic clones including the complete gene.  
CC CSG nucleic acids can be used to produce CSG polypeptides (see also  
W06545-53) in transformed host cells, as probes to detect disorders  
of the colon, partic. colon cancer and colon cancer metastasis, and  
in gene therapy.  
SQ Sequence 878 BP: 257 A; 179 C; 188 G; 241 T;

```

OY 2352 TGGGATGATTTATGACATGAGACAGCTCAGCATATATCATTCGATAAGTACAGTAT 2411
DB 361 TGGGATGATTTATGACATGAGACAGCTCAGCATATATCATTCGATAAGTACAGTAT 420
OY 2412 TCTGTATCTCAGAGACAGATTCATGATATCTTCAAGTGAATGATCTCTCATCC 2471
DB 421 TCTGTATCTCAGAGACAGATTCATGATATCTTCAAGTGAATGATCTCTCATCC 480
OY 2472 AAAGGAGCCACTCTGAGAGAGTCTTTTGTTTAAACAGAAACATTTACTTTGAAA 2531
DB 481 AAAGGAGCCACTCTGAGAGAGTCTTTTGTTTAAACAGAAACATTTACTTTGAAA 540
OY 2532 TGGCAGAGATCTTTTATGCTATTCAGGCTGTGATTAAGTCGATGATAAATCAGAAAT 2591
DB 541 TGGCAGAGATCTTTTATGCTATTCAGGCTGTGATTAAGTCGATGATAAATCAGAAAT 600
OY 2592 ATCCAACTTGCACGAGATATCTTTGTTATTCCTCCACAGACTCCGCGAGACACTAG 2651
DB 601 ATCCAACTTGCACGAGATATCTTTGTTATTCCTCCACAGACTCCGCGAGACACTAG 660
OY 2652 TCCTGATGAAGCTCTGCTCTTGT 2676
DB 661 TCCTGATGAAGCTCTGCTCTTGT 685

RESULT 2
VI6672
ID VI6672 standard; cDNA; 878 BP.
AC VI6672.
DE 22-JUN-1998 (first entry)
DE Polynucleotide sequence of a colon-specific gene.
KW Colon-specific gene; probe; detection; expression; human;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 2..685
FT /tag= a
FT /note= "no stop codon given"

US5733748-A.
PD 31-MAR-1998.
PF 06-JUN-1995; 469667.
PR 06-JUN-1995; US-469667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen C, Yu G;
P1 WPI: 98-229823/20.
DR P-PSDB; W46879.
DE Colon-specific nucleic acids - useful as probes for detecting colon
PT cancer micrometastases
PS Claim 15; Fig 5A-B; 51pp; English.
CC VI6668-81 represent polynucleotide sequences of partial or full-length
CC cDNA clones of colon-specific genes. The polynucleotides can be used
CC as probes to detect expression of the corresponding human genes, e.g. in
CC diagnostic assays for detecting micrometastases of colon cancer.
CC Recombinant cells containing the polynucleotides can be used to
CC produce the polypeptides, in order that antibodies can be raised and
CC used in further screening or diagnostics.
SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 24.4%; Score 685; DB 1; Length 878;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1992 TGTCTACTCAGATGATTTACAACTTATGACAGCATGATGATACGTTAAAGTCG 2051
DB 1 TGTCTACTCAGATGATTTACAACTTATGACAGCATGATGATGATGATGATGATGATG 60
OY 2052 GGCCTGGAGAGATTTAAAGCAGCAGAGAGTGTATACCCGCGAGAGTGTGAGCACT 2111
DB 61 GGCCTGGAGAGATTTAAAGCAGCAGAGAGTGTATACCCGCGAGAGTGTGAGCACT 120
OY 2112 GTACATACCTCTCTGAGATTGAGATGAAATACATGGAATCCACAGACCTGAAAT 2171

```

```

DB 121 GTACATACCTGCTGATGATGAGATGATGAAATACATGGAATCCACAGACCTGAAAT 180
OY 2172 TAATAGAGATGATGTTTCAACACAGCAAGTGTGTTTACAGAGACATCTCGGAGGCTC 2231
DB 181 TAATAGAGATGATGTTTCAACACAGCAAGTGTGTTTACAGAGACATCTCGGAGGCTC 240
OY 2232 ATTTGGCTTCTGATGTTTCCCAATATGCTCCCATACCTATCTCTTCCACCTGGCAAT 2291
DB 241 ATTTGGCTTCTGATGTTTCCCAATATGCTCCCATACCTATCTCTTCCACCTGGCAAT 300
OY 2292 CACCACTGAAAGGCGAAATTTACGCGGCGAGTCTCATTAATGATGATGATGATGATG 2351
DB 301 CACCACTGAAAGGCGAAATTTACGCGGCGAGTCTCATTAATGATGATGATGATGATG 360
OY 2352 TGGGATGATTTATGACATGAGACAGCTCAGCAAGTATATCATTTGAAATAGTACAGTAT 2411
DB 361 TGGGATGATTTATGACATGAGACAGCTCAGCAAGTATATCATTTGAAATAGTACAGTAT 420
OY 2412 TCTGTATCTCAGAGACAGATTCATGATATCTTCAAGTGAATGATCTCTCATCC 2471
DB 421 TCTGTATCTCAGAGACAGATTCATGATATCTTCAAGTGAATGATCTCTCATCC 480
OY 2472 AAAGGAGCCACTCTGAGAGAGTCTTTTGTTTAAACAGAAACATTTACTTTGAAA 2531
DB 481 AAAGGAGCCACTCTGAGAGAGTCTTTTGTTTAAACAGAAACATTTACTTTGAAA 540
OY 2532 TGGCAGAGATCTTTTATGCTATTCAGGCTGTGATTAAGTCGATGATAAATCAGAAAT 2591
DB 541 TGGCAGAGATCTTTTATGCTATTCAGGCTGTGATTAAGTCGATGATAAATCAGAAAT 600
OY 2592 ATCCAACTTGCACGAGATATCTTTGTTATTCCTCCACAGACTCCGCGAGACACTAG 2651
DB 601 ATCCAACTTGCACGAGATATCTTTGTTATTCCTCCACAGACTCCGCGAGACACTAG 660
OY 2652 TCCTGATGAAGCTCTGCTCTTGT 2676
DB 661 TCCTGATGAAGCTCTGCTCTTGT 685

RESULT 3
T22483
ID T22483 standard; cDNA to mRNA; 255 BP.
AC T22483.
DE 22-AUG-1996 (first entry)
DE Human gene signature HMG504094.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
P1 WPI: 95-206931/27.
DE Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1138-1139; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.

```

CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 255 BP; 76 A; 52 C; 39 G; 80 T;

Query Match 5.8%; Score 164; DB 1; Length 255;  
Best Local Similarity 100.0%; Pred. No. 2.3e-70;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2623 CCGCCAGAGACTCCGAGAGACCTAGTCGATGAAGTGGTCTGCTGCTAAT 2682  
DB 49 CTTCCACAGACTCCGAGAGACCTAGTCGATGAAGTGGTCTGCTGCTAAT 108  
QY 2683 ATTCATATCAACAGACACCATTCCTGCGATTTCATTTTAAATATGGAAGTGATA 2742  
DB 109 ATTCATATCAACAGACACCATTCCTGCGATTTCATTTTAAATATGGAAGTGATA 168  
QY 2743 GGAGAACTGAGCTGCTCATATAGCTAGGCTGAATTTTGTGAG 2786  
DB 169 GGAGAACTGAGCTGCTCATATAGCTAGGCTGAATTTTGTGAG 212

RESULT 4  
Q06308/c 0.7%; Score 19; DB 1; Length 2184;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ID 006308 standard; DNA: 710 BP.  
AC 006308;  
DT 29-JAN-1991 (first entry)  
DE Sequence of DNA fragment F7 of the human IFN-gamma receptor.  
KW IFN-gamma receptor; autoimmune disease; multiple sclerosis;  
KW hypersensitivity; ds.  
OS Homo sapiens.  
PN EP-393502-A.  
PD 24-OCT-1990.  
PF 11-APR-1990; 106992.  
PR 19-APR-1989; EP-810295.  
PA (HOFF ) HOFFMANN-LA ROCHE AG.  
PI Fountoulakis M, Garotta G, Stuber D;  
DR WPI: 90-322042/43.  
P-PSDB: R07472.  
PT Soluble interferon-gamma receptors - for treating auto-immune  
PT diseases, chronic inflammations, etc.  
PS Disclosure: Fig 18; 174pp; English.  
CC IFN-gamma is a therapeutically active agent in the treatment  
CC of autoimmune disease, allograft transplant rejections, multiple  
CC sclerosis, chronic inflammations and delayed hypersensitivity. It is  
CC also useful in identifying IFN-gamma agonists and antagonists.  
CC See also 006301.  
SQ Sequence 710 BP; 206 A; 147 C; 166 G; 191 T;

Query Match 0.7%; Score 19; DB 1; Length 710;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 TACAATATGATGAGAAAT 547  
DB 322 TACAATATGATGAGAAAT 304  
RESULT 5  
Q06301/c 0.7%; Score 19; DB 1; Length 2184 BP.  
ID 006301 standard; DNA: 2184 BP.  
AC 006301;  
DT 29-JAN-1991 (first entry)  
DE Sequence encoding Sac1/Asp181 fragment of plasmid pBABE carrying  
DE human Interferon gamma receptor gene.  
KW IFN-gamma receptor; autoimmune disease; multiple sclerosis;  
KW hypersensitivity; ds.  
OS Homo sapiens.  
FT Key Location/Qualifiers  
cds 85..1551

FT EP-393502-A. /\*tag- a  
PN 24-OCT-1990.  
PD 11-APR-1990; 106992.  
PF 19-APR-1989; EP-810295.  
PA (HOFF ) HOFFMANN-LA ROCHE AG.  
PI Fountoulakis M, Garotta G, Stuber D;  
DR WPI: 90-322042/43.  
P-PSDB: R07469.  
PT Soluble interferon-gamma receptors - for treating auto-immune  
PT diseases, chronic inflammations, etc.  
PS Disclosure: Fig 1; 174pp; English.  
CC Sequence may be used to transform prokaryotic or mammalian host  
CC cells via an expression vector, allowing production of the IFN-gamma  
CC receptor in pure form.  
CC The gene product is a therapeutically active agent in the treatment  
CC of autoimmune disease, allograft transplant rejections, multiple  
CC sclerosis, chronic inflammations and delayed hypersensitivity. It is  
CC also useful in identifying IFN-gamma agonists and antagonists.  
SQ Sequence 2184 BP; 688 A; 413 C; 451 G; 632 T;

Query Match 0.7%; Score 19; DB 1; Length 2184;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 TACAATATGATGAGAAAT 547  
DB 340 TACAATATGATGAGAAAT 322

RESULT 6  
V59671 0.7%; Score 19; DB 1; Length 2100 BP.  
ID V59671 standard; DNA: 2100 BP.  
AC V59671;  
DT 19-JAN-1999 (first entry)  
DE Human secreted protein gene 161 clone HAMB28.  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.  
PN M09839448-A2.  
PD 11-SEP-1998.  
PF 06-MAR-1998; U04493.  
PR 02-OCT-1997; US-061060.  
PR 07-MAR-1997; US-038621.  
PR 07-MAR-1997; US-040161.  
PR 07-MAR-1997; US-040162.  
PR 07-MAR-1997; US-040163.  
PR 07-MAR-1997; US-040333.  
PR 07-MAR-1997; US-040334.  
PR 07-MAR-1997; US-040336.  
PR 07-MAR-1997; US-040626.  
PR 11-APR-1997; US-043311.  
PR 11-APR-1997; US-043312.  
PR 11-APR-1997; US-043313.  
PR 11-APR-1997; US-043314.  
PR 11-APR-1997; US-043315.  
PR 11-APR-1997; US-043569.  
PR 11-APR-1997; US-043576.  
PR 11-APR-1997; US-043578.  
PR 11-APR-1997; US-043580.  
PR 11-APR-1997; US-043586.  
PR 11-APR-1997; US-043670.  
PR 11-APR-1997; US-043671.  
PR 11-APR-1997; US-043672.  
PR 11-APR-1997; US-043674.  
PR 23-MAY-1997; US-047492.



CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
SQ Sequence 447 BP; 120 A; 107 C; 95 G; 125 T;

Query Match 0.7%; Score 19; DB 1; Length 447;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 TACATATATGATGAGAAAT 547  
|||||  
DB 168 TACATATATGATGAGAAAT 186

RESULT 8  
V89358  
V89358 standard; cDNA; 299 BP.

AC V89358;  
DE EST clone C7493. (first entry)  
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
KW gene therapy; ss.  
OS Homo sapiens.  
PN WO9845436-A2.

PD 15-OCT-1998.  
PF 10-APR-1998; U06955.  
PR 10-APR-1997; US-838821.  
PA (GEM) GENETICS INST INC.  
PI Agostino M, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racie LA, Spaulding V, Treacy M,  
DR WPI: 99-070077/06.

PT New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.

PS Claim 1: Page 192: 618bp; English.  
CC The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
SQ Sequence 299 BP; 72 A; 65 C; 48 G; 114 T;

Query Match 0.7%; Score 19; DB 1; Length 299;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 563 GAAGATACAGCAGTAG 581  
|||||  
DB 193 GAAGATACAGCAGTAG 211

RESULT 9  
N50364/c  
N50364 standard; DNA; 3642 BP.  
AC N50364;  
DE 12-NOV-1991 (first entry)  
DE DNA fragment encoding herpes simplex virus 1 glycoprotein B.  
KW Herpes simplex virus 1; glycoprotein B; ss DNA; vaccine.  
OS Herpes simplex virus 1.  
FH Key Location/Qualifiers  
FT cds 438..3149  
FT /\*tag a  
PN MO8504587-A.

PD 24-OCT-1985.  
PF 04-APR-1985; U00587.  
PR 06-APR-1984; US-597784.  
PR 17-JUL-1984; US-631669.  
PI (CHIR-) Chiron Corp.  
PI Burke RL, Pachl C, Valenzuela PDT, Urdea MS;  
DR WPI: 85-276087/44.  
DR P-PSDB: P50312.

PT Recombinant herpes simplex vaccine - prep. by expression of DNA  
PT constructs in a eukaryotic host.  
PS Disclosure; Table 1 page 26-30; 80pp; English.  
CC Herpes simplex virus glycoprotein B or fragments may be used in a  
CC vaccine against HSV. Dosage is 10 micrograms to 2 mg/kg. The  
CC glycoprotein DNA is expressed in an eukaryotic host, esp.  
CC Saccharomyces cerevisiae, CHO cells and COS cells. Suitable plasmids  
CC are pYHS15, 116, 117, 118 and 119.  
SQ Sequence 3642 BP; 689 A; 1241 C; 1105 G; 607 T;

Query Match 0.6%; Score 18; DB 1; Length 3642;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 ACACCAAAAGATGCACAT 658  
|||||  
DB 3209 ACACCAAAAGATGCACAT 3192

RESULT 10  
N80908/c  
N80908 standard; DNA; 3474 BP.

AC N80908;  
DE 09-MAR-1992 (first entry)  
DE Sequence of Herpes simplex virus (HSV) glycoprotein B (gB) 2 DNA.  
KW Vaccine; herpes simplex virus; therapy; ss.  
OS Herpes simplex virus.  
FH Key Location/Qualifiers  
FT cds 309..3023  
FT /\*tag a

PN WO8802634-A.  
PD 21-APR-1988.  
PF 20-OCT-1987; U02709.  
PR 20-OCT-1986; US-921213.  
PR 20-JUL-1987; US-079605.  
PA (CHIR-) CHIRON CORP.  
PI Burke LR, Pachl C, Valenzuela P;  
DR WPI: 88-119368/17.  
DR P-PSDB: P80915.  
PT Vaccine for treatment of herpes simplex virus - contains  
PT recombinant HSV glyco-proteins B and D  
PS Example; Fig 4; 71pp; English.

CC Prepn. of recombinant gB and gD is described in WO88504587. The  
CC amino acid sequence and DNA sequence for gB 1 presented in P80914  
CC and N80907 differ from that originally presented in Table 1 of  
CC International Publication No. WO 85/04587. The DNA sequence in Table  
CC 1 contains an error in that an additional nucleotide (G) is listed  
CC at position 607 which resulted in a shift in reading frame relative  
CC to N80907 from which this nucleotide has been deleted.  
SQ Sequence 3474 BP; 666 A; 1183 C; 1058 G; 567 T;

Query Match 0.6%; Score 18; DB 1; Length 3474;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 ACACCAAAAGATGCACAT 658  
|||||  
DB 3086 ACACCAAAAGATGCACAT 3069

RESULT 11  
Q31988/c  
Q31988 standard; cDNA; 897 BP.

AC Q31988; 30-APR-1993 (first entry)  
 DE Guinea pig MBP-1 gene.  
 KM Major basic protein; anti-asthma; ss.  
 OS Cavia porcellus. Location/Qualifiers  
 FH Key cds  
 FT signal\_peptide /tag= a  
 FT signal\_peptide /tag= b  
 FT signal\_peptide /tag= c  
 FT mat\_peptide /note= "precursor peptide"  
 FT /tag= d  
 FT J04320686-A.  
 PN 11-NOV-1992.  
 PD 22-APR-1991: JP-090704.  
 PR 22-APR-1991: JP-090704.  
 PA (SAR) OTSUKA PHARM CO LTD.  
 WP1: 92-426880/52.  
 DR P-PSDB: R29716.  
 PT A guinea pig MBP gene - contains DNA base sequence coding MBP  
 PT gene having 1 of 2 specified aminoacid sequences  
 PS Claim 1: Page 15: 20pp; Japanese.  
 CC The sequence shows the cDNA sequence of the guinea pig major basic  
 CC protein (MBP)-1 gene. The sequence was obtained by first isolating MBP-1  
 CC and MBP-2 from the abdominal transudate of male guinea pigs treated  
 CC with polymyxin and ascaris extract. MBP-1 and MBP-2 were used to  
 CC design probes for screening of a guinea pig eosinophil cDNA library.  
 CC The MBPs may be prepd. in large amounts by this method and may be  
 CC used for the confirmation of the effect of an anti-asthma drug.  
 CC See also Q31989-93.  
 CC Sequence 897 BP; 201 A; 232 C; 282 G; 182 T;  
 SQ

Query Match 0.6%; Score 18; DB 1; Length 897;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 CAGGAAGCGATTATT 244  
 ||||||||||||||||  
 DB 876 CAGGAAGCGATTATT 859

RESULT 12  
 Q48497/c  
 ID Q48497 standard; DNA: 3472 BP.  
 AC Q48497; 29-MAR-1994 (first entry)  
 DT Glycoprotein B (GB2).  
 DE Glycoprotein; gB1; gB2; herpes simplex virus; HSV-1; HSV-2;  
 KM flanking; initiation; termination; transcription; translation;  
 KM regulatory sequence; ss.  
 OS Herpes simplex virus 2 strain 333.  
 FH Key cds  
 FT Location/Qualifiers  
 FT 309..3023  
 FT /tag= a  
 FT US5244792-A.  
 PN 14-SEP-1993.  
 PD 06-APR-1984; 597784.  
 PR 06-APR-1984; US-597784.  
 PR 20-OCT-1986; US-921730.  
 PR 20-SEP-1990; US-587179.  
 PA (CHIR) CHIRON CORP.  
 PI Burke RL, Pachl C, Valenzuela PDT;  
 WP1: 93-302641/38.  
 DR P-PSDB: R41779.  
 PT DNA construct for expressing HSV glycosylated polypeptide -  
 PT useful for vaccinating against HSV-1 and -2 infections  
 PS Disclosure; Fig 4; 33pp; English.  
 CC DNA constructs for expressing a glycosylated polypeptide in a  
 CC eukaryotic host cell comprises (1) an oligonucleotide sequence (OS)

CC free of natural flanking sequences, encoding glycoprotein B (GB)  
 CC of HSV, or its fragments, and (2) 5' initiating and 3' terminating  
 CC transcriptional and translational regulatory sequences flanking OS.  
 CC at least one of these sequences not being from HSV. The HSV-1 and  
 CC HSV-2 sequences are given in Q48496 and Q48497 respectively.  
 SQ Sequence 3472 BP; 661 A; 1193 C; 1053 G; 565 T;  
 SQ

Query Match 0.6%; Score 18; DB 1; Length 3472;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 ACACCAAAAGTGCACAT 658  
 ||||||||||||||||  
 DB 3083 ACACCAAAAGTGCACAT 3066

RESULT 13  
 Q73229/c  
 ID Q73229 standard; CDNA: 735 BP.  
 AC Q73229; 11-APR-1995 (first entry)  
 DE Soluble human interferon gamma receptor coding sequence.  
 KM Interferon; gamma; IFN; receptor; immunoglobulin; constant domain;  
 KM light chain; heavy chain; Ig; chimeric protein; fusion protein;  
 KM autoimmune disease; chronic inflammation; allotransplant; rejection;  
 KM multiple sclerosis; fulminant hepatitis; neurological disease; AIDS;  
 KM poliovirus; Lyme disease; septicemia; treatment; therapy;  
 OS delayed type hypersensitivity; ss.  
 FH Key cds  
 FT Location/Qualifiers  
 FT 1..735  
 FT /tag= a  
 FT /product= Soluble Interferon gamma receptor.  
 FT signal\_peptide 1..51  
 FT mat\_peptide /tag= b  
 FT /tag= c  
 FT EP-614981-A.  
 PN 14-SEP-1994.  
 PD 18-FEB-1993; 102452.  
 PR 05-MAR-1993; EP-810170.  
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.  
 PI Dembic Z, Garotta G, Gentz R;  
 WP1: 94-281208/35.  
 DR P-PSDB: R62023.  
 PT Chimeric human interferon-gamma receptor/immunoglobulin proteins  
 PT - used to inhibit binding of interferon-gamma to its specific  
 PT receptor in the treatment of illnesses  
 PS Disclosure; Figure 1; 29pp; English.  
 CC The soluble form of the interferon (IFN) gamma receptor comprises  
 CC the whole extracellular domain of the natural receptor from the N-  
 CC terminus to the transmembrane region, lacks the cytoplasmic and  
 CC transmembrane domains of the natural receptor and specifically binds  
 CC IFN-gamma. The sequence encoding the soluble IFN-gamma receptor can  
 CC be used in constructs encoding chimeric proteins where the other  
 CC component of the chimeric protein is part or all of the constant  
 CC domain of a human immunoglobulin heavy or light chain. The  
 CC recombinant proteins can be used to inhibit IFN-gamma binding to its  
 CC specific receptor. They can be used for the treatment of  
 CC illnesses, especially autoimmune diseases, chronic inflammation,  
 CC delayed type hypersensitivity, allotransplant rejections, multiple  
 CC sclerosis, fulminant hepatitis, inflammatory neurological diseases  
 CC and neurological complications of AIDS, poliovirus infections, Lyme  
 CC disease and septicemia. The presence of the immunoglobulin  
 CC component in the chimeric protein increases the proteins half life in  
 CC vivo.  
 SQ Sequence 735 BP; 223 A; 135 C; 170 G; 207 T;  
 SQ

Query Match 0.6%; Score 18; DB 1; Length 735;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 529 TACATATGATGAGAA 546  
 DB 256 TACATATGATGAGAA 239

RESULT 14  
 T43949

ID T43949 standard; DNA: 2040 BP.  
 AC T43949;  
 DT 18-AUG-1997 (first entry)  
 DE Sequence flanking marker 63-2 in HH region of chromosome 6p2.1.  
 KW Primer: polymerase chain reaction; amplify: hereditary haemochromatosis;  
 KW HH: mutation: HH-associated allele: base-pair polymorphism: HHP-1;  
 KW HHP-1; HHP-29; microsatellite repeat allele: genetic marker;  
 KW Interferon treatment; hepatitis C infection; ss.  
 OS Synthetic.  
 PN M09635803-A1.  
 PD 14-NOV-1996.  
 PE 08-MAY-1995; US-06583.  
 PR 08-MAY-1995; US-436074.  
 PR 15-NOV-1995; US-559302.  
 PR 09-FEB-1996; US-599252.  
 PA (MERC-) MERCATOR GENETICS INC.  
 PI Dryna DT. Feder JN, Gnlrke A, Kimmel BE, Thomas WJ;  
 PI Wolff RK.  
 DR WPI: 96-518691/51.  
 PT Diagnosing and genotyping of hereditary haemochromatosis (HH) -  
 PT using primers to detect specific polymorphisms of the HH gene on  
 PT chromosome 6p2.1 or novel microsatellite markers  
 PS Disclosure: Fig 1y: 67pp: English.  
 CC The sequences given in T43925-55 represent portions of the genome  
 CC surrounding several markers of the invention. The markers were  
 CC identified using the series of primer pairs given in T71901-72  
 CC which were used to determine the presence or absence of the common  
 CC hereditary haemochromatosis (HH) gene mutation in an individual. The  
 CC method comprised assessing genomic DNA from an individual for the  
 CC presence or absence of the HH-associated allele of the single base-pair  
 CC polymorphism HHP-1, HHP-19 or HHP-29, and/or at least one non-optional  
 CC marker comprising the following microsatellite repeat alleles of group  
 CC A and optionally of group B:  
 CC Group A: 19D9, 18B4, 1A2, 1E4, 24E2, 28B, 3321-1, 4073-1, 4440-1, 4440-2,  
 CC 731-1, 5091-1, 3216-1, 4072-2, 950-1, 950-2, 950-3, 950-4, 950-5, 950-6,  
 CC 950-8, 63-1, 63-2, 63-3, 65-1, 65-2, 373-8, 373-29, 68-1, 241-6, 241-29;  
 CC Group B: D6S464, D6S306, D6S258, D6S265, D6S109 and D6S1001.  
 CC The absence of the genotype indicates the likelihood of the presence of  
 CC the HH mutation. Knowledge of the new genetic markers allows the  
 CC definition of genotypes characteristic of heterozygous carriers and  
 CC homozygotes having a HH mutation in their genomic DNA. The potential for  
 CC HH in an individual interferes with the effectiveness of interferon  
 CC treatment for hepatitis C infection. By diagnosing this potential, the  
 CC responsiveness of interferon treatment may be evaluated.  
 CC Sequence 2040 BP; 713 A; 409 C; 398 G; 513 T;  
 SQ

Query Match 0.6%; Score 18; DB 1; Length 2040;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1647 CTTGTACTGCACAAA 1664  
 DB 195 CTTGTACTGCACAAA 212

RESULT 15  
 T69545  
 ID T69545 standard; DNA: 1386 BP.  
 AC T69545;  
 DT 26-AUG-1997 (first entry)  
 DE Rat pheromone receptor VNI cDNA.  
 KW Pheromone receptor; vomeronasal sensory neuron; social behaviour;  
 KW maternal behaviour; reproductive behaviour; fertility;  
 KW hormone secretion; ss.

OS Rattus sp.  
 FH Key Location/Qualifiers  
 FT cds 270..1217  
 FT /\*tag= a

PN M09714790-A1.  
 PD 24-APR-1997.  
 PE 18-OCT-1996; U16637.  
 PR 19-OCT-1995; US-005698.  
 PA (UYCO ) UNIV COLUMBIA NEW YORK.  
 PI Axel R, Dulac C;  
 DR WPI: 97-245107/22.  
 PT Nucleic acid molecule encoding vertebrate pheromone receptor -  
 PT useful to identify modulators for control of reproductive and social  
 PT behaviour, fertility and hormone secretion  
 PS Claim 1; Fig 8; 123pp: English.  
 CC A cDNA clone (T69545) codes for rat pheromone receptor VNI  
 CC (W19103), a member of a novel family of presumed 7-transmembrane  
 CC domain receptors that are evolutionarily independent of the odorant  
 CC receptors of the main olfactory epithelium (MOE). It was isolated  
 CC by differential screening of cDNA libraries constructed from rat  
 CC single vomeronasal sensory neurons. A VNI insert in pBluescript  
 CC has been deposited as ATCC 97294. VNI is expressed only in about  
 CC 4% of VSN and never in the MOE. cDNA clones for VN2-7 were obtd.  
 CC by PCR and hybridisation (see also T69546-50, VN2 sequence not  
 CC given). VN polypeptides (W19103-09) can be expressed in host-  
 CC vector systems for use in identifying modulators for control of  
 CC maternal, reproductive and social behavior, to increase fertility,  
 CC control hormone secretion and to regulate food uptake in humans and  
 CC animals. A human homologue, HG25 (T69551), has also been obtd.  
 SQ Sequence 1386 BP; 357 A; 322 C; 271 G; 435 T;

Query Match 0.6%; Score 18; DB 1; Length 1386;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1342 TCTGAGCTCAAGACTA 1359  
 DB 1042 TCTGAGCTCAAGACTA 1059

Search completed: August 5, 1999, 17:41:59  
 Job time: 6354 sec

This Page Blank (uspto)



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:43:13 ; Search time 3268.17 Seconds

(without alignments)  
103.150 Million cell updates/sec

Title: US-09-049-696-17

Perfect score: 106  
Sequence: 1 GGCAATTCACATTTTAAAAAT.....AAAATTAATCATTCATCCTT 106

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_htg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_htg1:\*  
35: gb\_htg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: gb\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	106	100.0	3311	11	AF039400	Homo sept

2	106	100.0	35278	11	AF039401	AF039401 Homo sapi
3	106	100.0	2826	42	AF127036	AF127036 Homo sapi
4	33	31.1	878	5	195746	195746 Sequence 8
5	20	18.9	143819	11	AC002994	AC002994 Homo sapi
6	19	17.9	16609	3	OCU63091	OCU63091 Oryctolagus
7	19	17.9	135686	9	HS47336	HS47336 Homo sapi
8	19	17.9	220218	11	AC006288	AC006288 Homo sapi
9	19	17.9	176564	34	HS135122	HS135122 Homo sapi
10	19	17.9	79760	34	HS501A4	HS501A4 Homo sapien
11	19	17.9	39220	36	CERT2A7	CERT2A7 Caenorhabd
12	19	17.9	76619	37	AC004722	AC004722 Drosophi1
13	19	17.9	47202	37	AC004724	AC004724 Drosophi1
14	18	17.0	2307	1	AF011388	AF011388 Brevibaci
15	18	17.0	4525	6	AB012111	AB012111 Clostridi
16	18	17.0	11584	6	AB012112	AB012112 Clostridi
17	18	17.0	12297	6	CBCBONT	X72793 Clostridium
18	18	17.0	82893	7	AB022218	AB022218 Arabidops
19	18	17.0	103200	9	HS30M3	AL031775 Human DNA
20	18	17.0	155661	10	HS399M14	296074 Human DNA s
21	18	17.0	158537	10	HSAC002069	AC002069 Human BAC
22	18	17.0	1263	10	HSKLT113A	X68757 H.sapiens K
23	18	17.0	4990	10	HSU73640	U73640 Human DNA s
24	18	17.0	36092	10	HSU73640	U73640 Human Chrom
25	18	17.0	197273	11	AC002461	AC002461 Human BAC
26	18	17.0	172241	11	AC004014	AC004014 Human BAC
27	18	17.0	74816	11	AC004855	AC004855 Homo sapi
28	18	17.0	157772	11	AC006212	AC006212 Homo sapi
29	18	17.0	92611	11	AC006398	AC006398 Homo sapi
30	18	17.0	163216	11	HUAC002289	AL078630 Homo sapi
31	18	17.0	154614	12	MM573K1	AC003021 Mus muscu
32	18	17.0	125536	34	AC003021	AL035467 Homo sapi
33	18	17.0	181663	34	HS288M22	AL035663 Homo sapi
34	18	17.0	105878	34	HS611N7	AL035460 Homo sapi
35	18	17.0	135613	34	HS860F19	AC004153 Plasmodiu
36	18	17.0	228625	35	AC004153	AC007353 Homo sapi
37	18	17.0	200877	35	AC007353	AC007353 Homo sapi
38	18	17.0	149804	35	AC007361	AC007377 Homo sapi
39	18	17.0	149494	35	AC007377	AC007361 Homo sapi
40	18	17.0	174369	35	AC007560	AC007560 Homo sapi
41	18	17.0	189036	42	AC006043	AC006043 Homo sapi
42	18	17.0	122697	42	AC006062	AC006062 Homo sapi
43	17	16.0	3048	1	CLOPERGASE	U04735 Clostridium
44	17	16.0	16649	2	LU93364	U93364 Lactococcus
45	17	16.0	827	36	TRKPEPTA	M33228 T.brucei K1

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS		AF039400				
DEFINITION		Homo sapiens calcium-dependent chloride channel-1 (hclcal) mRNA.				
ACCESSION		AF039400				
VERSION		g4009457				
KEYWORDS		AF039400.1 GI:4009457				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS		1 (bases 1 to 3311)				
TITLE		Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,R.D., Fuller,C.M. and Pauli,B.U.				
		Genomic cloning, molecular characterization, and functional analysis of human hclcal, the first human member of the family of Ca2+-activated Cl- channel proteins				
JOURNAL		Genomics 54 (2), 200-214 (1998)				
MEDLINE		99047526				
REFERENCE		2 (bases 1 to 3311)				
AUTHORS		Gruber,A.D., Elble,R. and Pauli,B.U.				
TITLE		Direct Submission				

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES  
SOURCE  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hclcal"  
352. .3096  
/gene="hclcal"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPKSSVFILLIHLLEGALSNSLIQLNNNGIEGIVADPNVP  
EDERLIQIKDMVTOASLYLEATGKREFFKNVAILPETWTKADYVRKLETVKNA  
DVLVAESTPENGDEPYEOMGNGEGERIHLPDFTAGKLAETGPOGKAFFHEWNA  
LRMGVDEYNDEKFEYLSNGRIQAVRCAGITGVNKKOOGSCYTRKCTENKVTGL  
YKGCPEVLOSROTERTASIMRAOHVDSIVFECTEONHNKEAPKRONKCLRSTEWYI  
RDEDFRKTTPMTTOPNPFTSLQIGORVCLVLDKSSGSMATGRNLRLNQAQFL  
LQTVELGSGWGMTFDSAHAVOSELQINSGRDYLAKRLPAASGGTISICGLNSA  
FTVIRKRYPTDGESEIVLTDGEDNTISGCFNEVQSAIHTVALGSAQELSELK  
MTGGLQTVASDQVQNNGLIDAFGALSNGAVSORSIQLESKGLTLQNSQWNGTVIY  
DSTVGKDTLELTWTTOPQIILMDPSGOGGFVVDNKKMAYLQIPGAKGTWY  
SLAASQTLITVTSRASNAITLPITVSTKNTKSPSPSLVYVANIROGASPIILA  
SVTALISVNGKITVLELDNGASADNTKDDGYSTRFTTIDNGRISVYRALGYN  
AARRVPIPOSGALYIPGWIENDEIQNNPPPEINKDDVOHKVCSRTSSGSPVAS  
DVPNAPIPDLFPPGQITDLNAETHGSLNLWTAPGDDYDHGTAKHYIIRISTSLD  
LRKFNSLQVNTALIPKEANSEVEFLKPENTTFENGTDLFAIOAVDKDLKSEI  
SNIRVSLFIPQTPPETPSPDETSAFCPIHINSTIPGHIILKIMKMWIGELQLSIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t

ORIGIN

Query Match 100.0%; Score 106; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 7.4e-44;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCAATTCACATTTTAAATATATGTGAAGTGATAGGAGAACTGCAGCTGCAATAGCC 60  
|||||  
Db 3034 GGCAATTCACATTTTAAATATATGTGAAGTGATAGGAGAACTGCAGCTGCAATAGCC 3093  
|||||

QY 61 TAGGCTGAATTTTTCAGATAAATTAATTAATCAATTCATCCTT 106  
|||||  
Db 3094 TAGGCTGAATTTTTCAGATAAATTAATTAATCAATTCATCCTT 3139  
|||||

RESULT 2  
AF039401 AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hclcal) gene,  
DEFINITION complete cds.  
ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreut,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human hclcal, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)

MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

FEATURES  
SOURCE  
1. .35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1. .1617  
/gene="hclcal"  
1. .33522  
/gene="hclcal"  
767. .1265  
/rpt\_family="L1 transposable element"  
/rpt\_type=dispersed  
1587. .1596  
/gene="hclcal"  
/note="putative"  
1618. .1708  
/gene="hclcal"  
/number=1  
join(1618. .1708, 1962. .2383, 6667. .6807, 6979. .7126, 9696. .9801, 15454. .15631, 18591. .18809, 19783. .20010, 22427. .24522, 24628, 26644. .26859, 27447. .27708, 28766. .28936, 31840. .32079, 32919. .33522)  
/product="calcium-dependent chloride channel-1"  
1962. .2383  
/gene="hclcal"  
/number=2  
join(2222. .2383, 6667. .6807, 6979. .7126, 9696. .9801, 15454. .15631, 18591. .18809, 19783. .20010, 22427. .24522, 24628, 26644. .26859, 27447. .27708, 28766. .28936, 31840. .32079, 32919. .33522)  
/gene="hclcal"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"

exon  
CDS  
exon  
TATA\_signal  
exon  
mRNA  
exon  
repeat\_region  
gene  
promoter  
TATA\_signal

exon  
exon  
exon  
exon  
exon  
exon

6667. .6807  
/gene="hclcal"  
/number=3  
6979. .7126  
/gene="hclcal"  
/number=4  
9696. .9801  
/gene="hclcal"  
/number=5  
15454. .15631  
/gene="hclcal"  
/number=6

```
exon      18591..18809
           /gene="hCLCA1"
           /number=7
exon      19783..20010
           /gene="hCLCA1"
           /number=8
exon      22253..22427
           /gene="hCLCA1"
           /number=9
exon      24522..24628
           /gene="hCLCA1"
           /number=10
exon      26644..26859
           /gene="hCLCA1"
           /number=11
exon      27447..27708
           /gene="hCLCA1"
           /number=12
exon      28766..28936
           /gene="hCLCA1"
           /number=13
exon      31840..32079
           /gene="hCLCA1"
           /number=14
exon      32919..33522
           /gene="hCLCA1"
           /number=15

BASE COUNT      11245 a      6870 c      7230 g      9933 t
ORIGIN

Query Match      100.0%; Score 106; DB 11; Length 35278;
Best Local Similarity 100.0%; Pred. No. 7.2e-44;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 GGCATTCACATTTTAAATATGATGGAAGTAGAGAGAGACAGCTGTCAATAGCC 60
         |||||||
Db      33248 GGCATTCACATTTTAAATATGATGGAAGTAGAGAGAGACAGCTGTCAATAGCC 33307

Oy      61 TAGGGCTGAATTTTGTGCAGATAAATAAATCAATCATTCCTT 106
         |||||||
Db      33308 TAGGGCTGAATTTTGTGCAGATAAATAAATCAATCATTCCTT 33353

RESULT      3
AF127036      2826 bp      mRNA      PRI      16-APR-1999
LOCUS      Homo sapiens calcium-activated chloride channel protein 1 (CACCL1)
DEFINITION      mRNA, complete cds.
ACCESSION      AF127036
NID      94585468
VERSION      AF127036.1 GI:4585468
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 2826)
AUTHORS      Agnol, M., Verma, T. and Culouscou, J.-M.
TITLE      Cloning of three human homologs of bovine epithelial chloride
channel
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 2826)
AUTHORS      Agnol, M. and Culouscou, J.-M.
TITLE      Direct Submission
JOURNAL      Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des
Carrières, Reuil-Malmaison 92500, France
FEATURES
Source      Location/Qualifiers
1..2826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue-type="small intestine; colon"
1..2826
gene
```

```
CDS
5..2749
/gene="CACCL1"
/gene="CACCL1"
/note="bovine epithelial chloride channel homolog"
/codon_start=1
/product="calcium-activated chloride channel protein 1"
/protein_id="A025487.1"
/db_xref="PIR:G458469"
/db_xref="GI:4585469"
/translation="MGPFKSSVFILHLLEGALNSLIOLNNNGEIVAVADPNVP
EDETLIQIDMTVOASLYLEAFGRKFEYKNVAILPEPTWKRADYRKLKYNA
DLYAESTPRGNDPEPYEOMKNGEGERHLPPDPAGKKLAEGPGRAFYENAH
LRNGVPEYNNDEKFTYLSNGRIQAVRSASITITNYVKKOGSCITRKTFNKGVL
YKGECEVLOSROTERKASIFEAQHVDSIVFEQDNKKAPKONKCNLSTWEYI
RDSDEFKRTTPTMTOPNPPTFSLQIGORIVCLVLDKSGMAIGNRLNLAGQLEFL
IQVLELSWVGMVTFDPAHVOSELQIINGSDRDYTLAKRLPAASGTSICGLRSA
FTYLRKKYPTDGSSEIVLLTDGENDTISGCENYKOSGAILHTVALGSAOLELELSK
MTGGLQRYASDOYONNGLIDAPGALSNGNVAORSIOLESKGLTIONSOMNGTVY
DSYVGKDTLLITTTTTPPOLIMLMDPSGAKOGSFYVDKNTKMAIQLIPGIKYGTVKI
SLQASSQTLITVTSRASNAFLPITVTSKNTDTSKFPSPLVVYANIROGASPTLKA
SVTALIESVNGKTVLELDNGAGADATKDDGVYSRFTYTDNGRYSVVRALGVN
AARRRVIPQOSGALYIPGMIENDEIOWNPRIKMDYOHKQVCSFSSGSFVAS
DVNPAPIPDLPFGOITDLKAEIHGSLILNLTWAPDDYDHGAHKYIIRISTSID
LRDKNESLQVNTTALIPKEANSEEVFLFPEVITFENGTDLFIAOAVDKVLSKI
SNARVSLFIPPTPPTPSPDETSAFCPIHINSTITPGIHILIKIMKWIQELSLA
"

BASE COUNT      875 a      623 c      632 g      696 t
ORIGIN

Query Match      100.0%; Score 106; DB 42; Length 2826;
Best Local Similarity 100.0%; Pred. No. 7.4e-44;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 GGCATTCACATTTTAAATATGATGGAAGTAGAGAGACAGCTGTCAATAGCC 60
         |||||||
Db      2687 GGCATTCACATTTTAAATATGATGGAAGTAGAGAGACAGCTGTCAATAGCC 2746

Oy      61 TAGGGCTGAATTTTGTGCAGATAAATAAATCAATCATTCCTT 106
         |||||||
Db      2747 TAGGGCTGAATTTTGTGCAGATAAATAAATCAATCATTCCTT 2792

RESULT      4
195746      878 bp      DNA      PAT      17-JUL-1998
LOCUS      Sequence 8 from patent US 5733748.
DEFINITION      195746
ACCESSION      93840216
NID      195746.1 GI:3940216
VERSION      195746.1
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 878)
AUTHORS      Yu, G. and Kosen, C.
TITLE      Colon specific genes and proteins
JOURNAL      Patent: US 5733748-A 8 31-MAR-1998;
FEATURES
Source      Location/Qualifiers
1..878
/organism="unknown"

BASE COUNT      257 a      179 c      188 g      241 t      13 others
ORIGIN

Query Match      31.1%; Score 33; DB 5; Length 878;
Best Local Similarity 100.0%; Pred. No. 4.7e-07;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 GGCATTCACATTTTAAATATGATGGAAGTAGG 33
         |||||||
Db      717 GGCATTCACATTTTAAATATGATGGAAGTAGG 749
```

RESULT 5  
AC002994/c  
LOCUS AC002994 143819 bp DNA PRI 10-NOV-1997  
DEFINITION Homo sapiens chromosome 17, clone HRC987K16, complete sequence.  
AC002994  
NID 92599233  
VERSION AC002994.1 GI:2599233  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Birren,B., Fasnman,K., McKernan,K., Munro,C., Nusbaum,C.,  
Richardson,P. and Lander,E.  
TITLE Homo sapiens chromosome 17, clone HRC987K16  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 143819)  
Hawkins,T.L., Birren,B.W., Fasnman,K.H., Nusbaum,C., Lander,E.S.,  
McKernan,K., Munro,C., Richardson,P., Baldwin,J., Barna,N.,  
Cantu,C., Chang,A., Cooke,P., Daly,M.J., Devon,K., Dewar,K.,  
Duretto,B., Forrest,C., Gage,D., Gensheimer,S., Geradigery,K.,  
Gilmartin,I., Hagos,B., Halphen,I., Harris,K., Howland,J.C.,  
Huang,J., Hul,L., Jacotoc,L., Kirby,A., Lane,M., Mackenzie,J.,  
Marquis,N., McDermott,J., Molla,M., Morrow,J., Nachman,A.,  
Naylor,J., Nusbaum,C., O'Connor,T., Olotu,A., Peterson,K.,  
Reeve,M.P., Roberts,D., Rollins,G., Sarnak,A., Shiu,P., Shyam,R.,  
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,  
Walmer,F., Zemtseva,I. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (06-OCT-1997) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE  
AUTHORS 3 (bases 1 to 143819)  
Birren,B., Fasnman,K., McKernan,K., Munro,C., Nusbaum,C.,  
Richardson,P., Lander,E., Baldwin,J., Barna,N., Cantu,C.,  
Cooke,P., Daly,M.J., Devon,K., Dewar,K., Duretto,B., Forrest,C.,  
Gage,D., Gensheimer,S., Geradigery,K., Gilmartin,I., Hagos,B.,  
Halphen,I., Harris,K., Howland,J.C., Huang,J., Hul,L., Jacotoc,L.,  
Kirby,A., Lane,M., Mackenzie,J., Marquis,N., McDermott,J.,  
Molla,M., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Olotu,A.,  
Peterson,K., Roberts,D., Rollins,G., Sarnak,A., Shiu,P., Shyam,R.,  
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,  
Zemtseva,I. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (10-NOV-1997) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Nov 10, 1997 this sequence version replaced gi:2598080.  
The Staden databases, finishing information, and all  
chromatographic files used in the assembly of this clone are  
available from our anonymous ftp site.  
All repeats were identified using RepeatMasker: Smit, A.F.A. &  
Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html.  
FEATURES  
SOURCE  
1. 143819  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HRC987K16"  
/clone\_id="Peter de Jong/ human PAC library"  
/map="17"  
/chromosome="17"  
complement(404..452)  
/rpt\_family="CAAAA)n"  
complement(455..706)  
/rpt\_family="Alusx"  
complement(1273..1348)  
/rpt\_family="AT-rich"  
1440..1472  
/rpt\_family="AT-rich"  
2856..2945  
/rpt\_family="AT-rich"  
/rpt\_family="CAAAA)n"

repeat\_region 2956..2995  
/rpt\_family="MER58B"  
repeat\_region 2996..3094  
/rpt\_family="MER44A"  
repeat\_region 3093..3314  
/rpt\_family="MER44B"  
repeat\_region 3339..3385  
/rpt\_family="MER58B"  
repeat\_region 3420..3658  
/rpt\_family="Alusx"  
repeat\_region 3659..3748  
/rpt\_family="MER58B"  
repeat\_region 4460..4484  
/rpt\_family="AT-rich"  
5284..5559  
/rpt\_family="L1"  
5573..6010  
/rpt\_family="L1"  
6024..6576  
/rpt\_family="L1PA16"  
6530..7043  
/rpt\_family="L1"  
7025..7269  
/rpt\_family="L1"  
7292..7647  
/rpt\_family="L1"  
7644..8145  
/rpt\_family="L1PA15"  
8147..8445  
/rpt\_family="Alusx"  
8466..9239  
/rpt\_family="L1PA15"  
9240..9265  
/rpt\_family="CAA)n"  
9662..10379  
/rpt\_family="L1M4"  
10482..11290  
/rpt\_family="L1M4"  
11348..11888  
/rpt\_family="MER68A"  
12043..12502  
/rpt\_family="L1M7"  
12512..12796  
/rpt\_family="AluJb"  
12797..13383  
/rpt\_family="L1M7"  
13403..13695  
/rpt\_family="Alusx"  
13701..14132  
/rpt\_family="L1M7"  
complement(14152..14433)  
/rpt\_family="Alusg"  
complement(14505..14804)  
/rpt\_family="AluY"  
complement(14805..14961)  
/rpt\_family="Alusg/x"  
14962..17391  
/rpt\_family="L1M7"  
17444..17660  
/rpt\_family="L1"  
complement(18127..19262)  
/rpt\_family="L1M7"  
complement(19263..19388)  
/rpt\_family="MSTB"  
complement(19384..19511)  
/rpt\_family="MSTB"  
complement(19528..19747)  
/rpt\_family="L1"  
19756..20043  
/rpt\_family="Alusg"  
complement(20044..20226)  
/rpt\_family="L1"  
complement(20269..20580)

```

repeat_region /rpt_family="LIPAS" complement(20571. .20905)
repeat_region /rpt_family="LI" complement(21290. .21487)
repeat_region /rpt_family="LIM3/4" complement(21489. .21779)
repeat_region /rpt_family="ALuNb" 22187. .22464
repeat_region /rpt_family="AluSx" 22466. .22519
repeat_region /rpt_family="(TAAA)n" complement(22524. .22894)
repeat_region /rpt_family="LIM3/4" complement(23015. .23315)
repeat_region /rpt_family="AluNb" 23880. .23930
repeat_region /rpt_family="AT_rich" complement(24251. .24432)
repeat_region /rpt_family="AluYo" 24508. .24553
repeat_region /rpt_family="AT_rich" 24750. .24794
repeat_region /rpt_family="AT_rich" 25881. .25939
repeat_region /rpt_family="MIR" 26160. .26208
repeat_region /rpt_family="AT_rich" complement(26220. .26455)
repeat_region /rpt_family="MER58A" 26500. .26750
repeat_region /rpt_family="AluSq" 26755. .26856
repeat_region /rpt_family="AT_rich" 26892. .26981
repeat_region /rpt_family="MIR" complement(28359. .28657)
repeat_region /rpt_family="AluSq" 30006. .30096
repeat_region /rpt_family="MIR" 30278. .30635
repeat_region /rpt_family="THE1C" 30730. .31030
repeat_region /rpt_family="AluY" complement(31142. .31447)
repeat_region /rpt_family="AluSx" 31723. .31781
repeat_region /rpt_family="MIR" complement(33281. .33320)
repeat_region /rpt_family="AT_rich" complement(33477. .33531)
repeat_region /rpt_family="(GAAA)n" complement(33543. .34783)
repeat_region /rpt_family="LIPAZ" complement(34803. .38450)
repeat_region /rpt_family="LI" 38742. .38872
repeat_region /rpt_family="FLAM_A"

```

Query Match	18.9%	Score 20:	DB 11,	Length 143819;
Best Local Similarity	100.0%	Pred. No. 1.6;		
Matches 20:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps
QY	11	TTTTAAATAATTATGTGGAG	30	
DB 107984		TTTTAAATAATTATGTGGAG	107965	
RESULT	6			
LOCUS	OCU63091			
DEFINITION	OCU63091 16609 bp DNA	MM	10-FEB-1997	
	Oryctolagus cuniculus beta-globin-like gene cluster locus control			
	region.			
ACCESSION	U63091	L05833	L05835	

NID	g1835272
VERSION	063091.1
KEYWORDS	GI:1835272
SOURCE	Oryctolagus cuniculus.
ORGANISM	Oryctolagus cuniculus Eukaryota; Lagomorpha; Chordata; Craniata; Vertebrata; Mammalia; Euarchonta; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE	1 (bases 2858 to 4174, 6636 to 13233) Hardison,R., Xu,J., Jackson,U., Mansberger,J., Selifonova,O., Grotch,B., Bliesecker,J., Petykowska,H. and Miller,W. Comparative analysis of the locus control region of the rabbit beta-like gene cluster: H3 increases transient expression of an embryonic epsilon-globin gene Nucleic Acids Res. 21 (5), 1265-1272 (1993)
AUTHORS	93219091
JOURNAL	2 (bases 1 to 16609) Sligfom,J.L., Bock,J.H., Tagle,D.A., Gumucio,D.L., Goodman,M., Stojanovic,N., Jackson,J., Miller,W. and Hardison,R. The complete sequences of the galago and rabbit beta-globin locus control regions: extended sequence and functional conservation outside the cores of phase hypersensitive sites Genomics 39 (1), 90-94 (1997)
MEDLINE	97179204
REFERENCE	3 (bases 1 to 16609) Sligfom,J.L., Bock,J.H., Tagle,D.A., Gumucio,D.L., Goodman,M., Stojanovic,N., Jackson,J.D., Miller,W. and Hardison,R. Direct Submission Submitted (05-JUL-1996) Biochemistry and Molecular Biology, The Pennsylvania State University, 206 Althouse Laboratory, University Park, PA 16802, USA
COMMENT	On Feb 10, 1997 this sequence version replaced gi:165033 gi:165034
FEATURES	Location/Qualifiers
source	1..16609 /organism="Oryctolagus cuniculus" /db_xref="taxon:9986" /clone="lambdaRbetag3.1 and lambdaRbetag8.3" /chromosome="1" /map="ig14-21" complement(1111..>1503) /note="ORF" /codon_start=1 /protein_id="AAC48696.1" /db_xref="PID:g1835273" /db_xref="GI:1835273" TNNVFFLTDIPVEVRHAIFFHFVF" complement(2135..2433) /note="Rabbit Lloc repeat, homologous to human LHS repeat." 2807..3013 /note="Insert in rabbit (not in human or galago)" 3337..3627 /note="Homolog to hypersensitive site 4 of beta-globin locus control region" 3419..3423 /standard_name="CDP" /note="putative" /bound_moiety="CCAAT-displacement factor" 3512..3518 /note="putative" /bound_moiety="AP1" 3567..3572 /note="putative" /bound_moiety="GARAI" 4355..4683 /note="rabbit C repeat, analogous to human A1u repeat" 4699..5038 /note="rabbit C repeat, analogous to human A1u repeat" 5792..5864 /note="insert in rabbit (not in human, galago, mouse)" 7815..7823 /note="putative"

	misc_feature	/bound_molety="NFE2/AP1"	7853. .8029	/note-"Homolog to hypersensitive site 3 of beta-globin locus control region"
	protein_bind	/note-"putative"	7856. .7861	/bound_molety="GATA1"
	protein_bind	/note-"putative"	7900. .7905	/bound_molety="GATAT"
	protein_bind	/note-"putative"	7937. .7942	/bound_molety="GATAT"
	protein_bind	/note-"putative"	7940. .7945	/bound_molety="GATAT"
	protein_bind	/note-"putative"	7968. .7971	/bound_molety="CAC-BP"
	protein_bind	/note-"putative"	7992. .7997	/bound_molety="GATAT"
	repeat_region	/note-"rabbt C repeat, analogous to human Alu repeat"	10203. .10556	/note-"rabbt C repeat, analogous to human Alu repeat"
	enhancer	/note-"Homolog to hypersensitive site 2 of beta-globin locus control region"	11452. .11826	/note-"rabbt C repeat, analogous to human Alu repeat"
	protein_bind	/note-"putative"	11620. .11628	/bound_molety="NFE2/AP1"
	protein_bind	/note-"putative"	11630. .11638	/bound_molety="NFE2/AP1"
	protein_bind	/note-"putative"	11640. .11648	/bound_molety="NFE2/AP1"
	protein_bind	/note-"putative"	11670. .11675	/standard_name="E box"
	protein_bind	/note-"putative"	11699. .11704	/bound_molety="unknown"
	repeat_region	/note-"rabbt C repeat, analogous to human Alu repeat"	12950. .13253	/note-"rabbt C repeat, analogous to human Alu repeat"
	repeat_region	/note-"rabbt C repeat, analogous to human Alu repeat"	14754. .15053	/note-"rabbt C repeat, analogous to human Alu repeat"
	repeat_region	/note-"rabbt C repeat, analogous to human Alu repeat"	16149. .16609	/note-"Homolog to hypersensitive site 1 of beta-globin locus control region"
	misc_feature	/note-"Homolog to hypersensitive site 1 of beta-globin locus control region"	3299 c 3171 g 5067 t	2 others
	BASE COUNT	5070 a	3299 c	3171 g 5067 t
	ORIGIN			
	Query Match	17.9%; Score 19; DB 3; Length 16609;		
	Best Local Similarity	100.0%; Pred.No. 5.1;		
	Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0.		
OY	81 ATAAATAAAAATCAATT 99			
DB	5004 AYAATATAAATCAATT 5022			
RESULT	7			
LOCUS	HS473J6	135686 bp	DNA	PRI 12-MAY-1997
DEFINITION	Human DNA sequence from PAC 473J6 on chromosome X contains STS.			
ACCESSION	Z83827			
NID	G2094790			
VERSION	Z83827.1	GI:2094790		
KEYWORDS	X.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			

```

REFERENCE      Primates; Catarrhini; Hominiidae; Homo.
AUTHORS        1 (bases 1 to 135686)
TITLE          WALKINSON, J
JOURNAL        Direct Submission
               Submitted (01-MAY-1997) Sanger Centre, Hinxton, Cambridgeshire,
               CB3 0SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
               requests: clonerequests@sanger.ac.uk
               ON MAY 14, 1997 this sequence version replaced g1:1772969.
COMMENT         IMPORTANT: This sequence is the entire insert of clone 473j6. This
               sequence was generated from part of bacterial clone contigs of
               human chromosome X, constructed by the Sanger Centre chromosome X
               mapping group. Further information can be found at
               http://www.sanger.ac.uk/HGP/chrX/
               This sequence has been finished according to sequence map criteria
               as follows. An attempt is made to resolve all sequencing problems,
               such as compressions and repeats, but not necessarily within known
               annotated human repeat sequence elements (e.g. Alu). Where the
               sequence is ambiguous, there is an annotation using the 'unsure'
               feature key.
               The true left end of clone 473j6 is at 1 in this sequence. The true
               right end of clone 473j6 is at 135686.
               473j6 is from the library RPCR3 constructed at the Roswell Park
               Cancer Institute by the group of Pieter de Jong.
               For further details see http://bacpac.med.buffalo.edu/.

FEATURES
Source
Location/Qualifiers
1..135686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="473j6"
/clone_1tb="RPCR1-3"
1..93
/note="MSTA repeat: matches 93. .1 of consensus"
repeat_region
558..764
/note="MER25 repeat: matches 1504. .1297 of consensus"
repeat_region
1336..1555
/note="MER25 repeat: matches 1296. .1079 of consensus"
repeat_region
1584..1877
/note="AlusX repeat: matches 297. .1 of consensus"
repeat_region
1889..2345
/note="MER25 repeat: matches 1057. .602 of consensus"
repeat_region
2533..3027
/note="MER25 repeat: matches 548. .49 of consensus"
repeat_region
3476..3594
/note="FLAN_A repeat: matches 128. .11 of consensus"
repeat_region
3736..4786
/note="THEIB-INTERNAL repeat: matches 107. .1167 of
consensus"
4790..5109
/note="THEIC repeat: matches 54. .368 of consensus"
repeat_region
5113..6250
/note="L1 repeat: matches 3226. .2078 of consensus"
repeat_region
6241..6480
/note="L1MB7 repeat: matches 242. .2 of consensus"
repeat_region
6333..7040
/note="L1 repeat: matches 5390. .4668 of consensus"
repeat_region
7044..7347
/note="AlusQ repeat: matches 1. .303 of consensus"
repeat_region
7370..9247
/note="L1 repeat: matches 4680. .2773 of consensus"
repeat_region
9296..9607
/note="Aluo repeat: matches 1. .302 of consensus"
repeat_region
9631..9663
/note="AlusC repeat: matches 36. .4 of consensus;
incomplete repeat"
repeat_region
9667..10620
/note="L1 repeat: matches 3978. .3022 of consensus"
repeat_region
10657..10951
/note="Aluo repeat: matches 301. .12 of consensus"
repeat_region
11983..12263
/note="Aluo repeat: matches 3. .277 of consensus;
incomplete repeat"

```

```
repeat_region 13944..13983
/note="20 copies of 2 mer 90 & conserved"
repeat_region 13984..14273
/note="AluYo repeat: matches 293. .6 of consensus"
repeat_region 16530..16701
/note="MLT1F repeat: matches 348. .539 of consensus"
repeat_region 17006..17206
/note="MLT1F repeat: matches 541. .343 of consensus"
repeat_region 17212..17582
/note="THEIC repeat: matches 371. .1 of consensus"
repeat_region 17588..19176
/note="THE1B-INTERNAL repeat: matches 1574. .1 of
consensus"
repeat_region 19179..19547
/note="THEIC repeat: matches 369. .2 of consensus"
repeat_region 19552..19776
/note="MLT1F repeat: matches 339. .113 of consensus"
repeat_region 23101..23301
/note="MLT1-INTERNAL repeat: matches 494. .287 of
consensus"
repeat_region 23368..23477
/note="MER34 repeat: matches 506. .401 of consensus"
repeat_region 24098..24307
/note="MLT1F repeat: matches 480. .268 of consensus"
repeat_region 24303..24727
/note="MER39 repeat: matches 6. .409 of consensus"
repeat_region 24813..24974
/note="MLT1F repeat: matches 272. .98 of consensus"
repeat_region 24893..24977
/note="MLT1G repeat: matches 186. .101 of consensus"
repeat_region 25312..26884
/note="LI repeat: matches 5183. .3613 of consensus"
repeat_region 26886..27112
/note="LI repeat: matches 5166. .5390 of consensus"
repeat_region 26962..27853
/note="LIpA2 repeat: matches 1. .891 of consensus"
repeat_region 29189..29318
/note="FLAM_C repeat: matches 1. .130 of consensus"
repeat_region 31274..31580
/note="AluSg repeat: matches 303. .1 of consensus"
repeat_region 32974..33269
/note="AluSc repeat: matches 1. .295 of consensus"
repeat_region 33593..33915
/note="MER44A repeat: matches 2. .315 of consensus"
repeat_region 33690..33911
/note="MER44B repeat: matches 283. .514 of consensus"
repeat_region 34675..34762
/note="MER34 repeat: matches 151. .56 of consensus"
repeat_region 35890..35972
/note="MER4B repeat: matches 539. .457 of consensus"
repeat_region 35973..36271
/note="LTR8 repeat: matches 691. .388 of consensus"
repeat_region 36364..36430
/note="LTR8 repeat: matches 317. .251 of consensus"
repeat_region 36562..36719
/note="LTR8 repeat: matches 165. .1 of consensus"
repeat_region 36983..37138
/note="LTR8 repeat: matches 146. .1 of consensus"
repeat_region 37139..37276
/note="MER4B repeat: matches 460. .323 of consensus"
repeat_region 37677..37887
/note="MER4B repeat: matches 213. .1 of consensus"
repeat_region 37978..38325
/note="match: STS L41748"
repeat_region 42104..42309
/note="LIpA5 repeat: matches 685. .890 of consensus"
repeat_region 49491..49516
/note="13 copies of 2 mer 92 & conserved"
repeat_region 49686..50316
/note="LI repeat: matches 4487. .5147 of consensus"
repeat_region 50306..52141
/note="LI repeat: matches 3536. .5390 of consensus"
repeat_region 52002..52648
/note="LIpA6 repeat: matches 3. .675 of consensus"
repeat_region 52644..53044
/note="MLT2B repeat: matches 1. .404 of consensus"
repeat_region 53047..53090
/note="22 copies of 2 mer 82 & conserved"
repeat_region 53141..53215
/note="MLT2G repeat: matches 317. .388 of consensus"
repeat_region 53159..53215
/note="MLT2B repeat: matches 390. .444 of consensus"
repeat_region 53217..53414
/note="MLT2 internal repeat: matches 2. .199 of consensus"
repeat_region 53415..53690
/note="AluX repeat: matches 19. .292 of consensus;
incomplete repeat"
repeat_region 53964..54006
/note="MADE1 repeat: matches 38. .80 of consensus"
repeat_region 55614..56278
/note="LIpA7 repeat: matches 890. .212 of consensus"
repeat_region 56278..56700
/note="LI repeat: matches 3941. .4388 of consensus"
repeat_region 57595..57763
/note="AluYo repeat: matches 302. .138 of consensus;
incomplete repeat"
repeat_region 58184..58515
/note="MLT2 internal repeat: matches 2057. .2400 of
consensus"
repeat_region 58676..59124
/note="MLT2 internal repeat: matches 3467. .3925 of
consensus"
repeat_region 59158..59555
/note="MLT1B repeat: matches 1. .390 of consensus"
repeat_region 59536..59638
/note="MLT1-INTERNAL repeat: matches 1. .84 of consensus"
repeat_region 60688..60871
/note="AluYb repeat: matches 7. .178 of consensus;
incomplete repeat"
repeat_region 60898..61206
/note="AluX repeat: matches 302. .1 of consensus"
repeat_region 61242..61373
/note="AluYo repeat: matches 169. .301 of consensus;
incomplete repeat"
repeat_region 63678..63849
/note="LIpA3 repeat: matches 44. .225 of consensus"
repeat_region 64020..64331
/note="LI repeat: matches 1737. .1426 of consensus"

Query Match 17.9%; Score 19; DB 9; Length 135686;
Best Local Similarity 100.08; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 TCAGATTAATAATAAT 95
Db 31748 TCAGATTAATAATAAT 31766

RESULT 8
AC006288 AC006288 220218 bp DNA PRI 16-JAN-1999
LOCUS Homo sapiens chromosome 9, clone hRPK.295_D_22, complete sequence.
DEFINITION AC006288
ACCESSION 94160142
VERSION 94160142
KEYWORDS AC006288.1 GI:4160142
SOURCE HTG.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 220218)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 9, clone hRPK.295_D_22
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 220218)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
```

```

Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deakre,K., Depayre,E., Devon,K., Dewar,K.,
Donegan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A.,
Heras,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A.,
Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,
McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C.,
Sudramanjan,A., Testaye,S., Torturella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (05-JAN-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 220218)
REFERENCE
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deakre,K., Depayre,E., Devon,K., Dewar,K.,
Donegan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Hearford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Sudramanjan,A.,
Testaye,S., Torturella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (16-JAN-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 16, 1999 this sequence version replaced gi:4156127.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
SOURCE
1. 220218
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK.295_D_22"
/clone_11b="RPCI-11 human BAC library"
/map="9"
/chromosome="9"
/complement(110..207)
repeat_region
/rpt_family="L2"
229..361
/rpt_family="L2"
complement(708..721)
/notes="Single-stranded coverage."
repeat_region
complement(838..908)
/rpt_family="MIR"
1338..1480
/rpt_family="L1PA13"
1481..1784
/rpt_family="AluX"
1762..1766
/notes="Single-stranded coverage."
repeat_region
1785..1848
/rpt_family="L1PA13"
3254..3294
/rpt_family="(TG)n"
complement(3616..3774)
repeat_region
/rpt_family="MIR"
complement(3869..3989)
repeat_region
/rpt_family="MIR"
5954..5983
/rpt_family="AT-rich"
6377..6685
/rpt_family="AluB"
repeat_region
7893..7995
/rpt_family="MIR"
complement(7993..8191)
repeat_region
/rpt_family="MIR"
8210..8287
/rpt_family="MIR"
8335..8816
/rpt_family="AluY"
9441..9733
/rpt_family="AluY"
10276..10442
/rpt_family="MERSA"
10506..10768
/rpt_family="L1PA10"
complement(10791..10927)
repeat_region
/rpt_family="L1PB3"
11357..11444
/rpt_family="L2"
complement(11512..11814)
repeat_region
/rpt_family="AluB"
complement(12044..13284)
/rpt_family="L1MA3"
13480..13551
/rpt_family="(CA)n"
1466..14910
/rpt_family="L1PA7"
15452..15994
/rpt_family="MLR2A"
16422..16523
/rpt_family="L2"
complement(16692..16997)
/rpt_family="AluX"
17446..17491
/rpt_family="GA-rich"
complement(19930..20046)
/rpt_family="L2"
complement(20196..20623)
/rpt_family="L2"
complement(21523..21744)
/rpt_family="MIR"
complement(23039..23125)
/rpt_family="MIR"
complement(23217..23353)
/rpt_family="L1MA6"
23614..24959
/rpt_family="L2"
24960..25270
/rpt_family="AluSg"
25271..25855
/rpt_family="L2"
25957..26136
/rpt_family="L2"
27312..27410
/rpt_family="MIR"
28115..28198
/rpt_family="L2"
28313..28350
/rpt_family="MIR"
complement(28351..28664)
/rpt_family="MERS3"
28665..28724
/rpt_family="MIR"
28723..28789
/rpt_family="L2"
29101..29226
/rpt_family="MIR"
30024..30047
/rpt_family="AT-rich"
30930..31159
/rpt_family="MIR"
31265..31274
/notes="Single-stranded coverage."
complement(31820..32122)

```



```

repeat_region /rpt_family="AluY"
33342..33419 /rpt_family="AT-rich"
repeat_region complement(33510..33615)
/rpt_family="MIR"
repeat_region 33936..34265
/rpt_family="LTR37A"
repeat_region 34665..34740
/rpt_family="MIR"
repeat_region complement(35242..35487)
/rpt_family="AluY"
repeat_region complement(35757..35843)
/rpt_family="L2"
repeat_region 36202..36426
/rpt_family="MIR"
repeat_region 36509..36570
/rpt_family="MIR"
repeat_region 36707..36966
/rpt_family="MIR"
repeat_region 37262..37322
/rpt_family="T-rich"
repeat_region complement(37885..37989)
/rpt_family="MIR"
repeat_region complement(37990..38308)
/rpt_family="MT1H"
repeat_region 38345..38390
/rpt_family="GA-rich"
repeat_region complement(38424..38580)
/rpt_family="MT1H"
repeat_region 38583..38606
/rpt_family="(TTG)n"
repeat_region complement(39574..39888)
/rpt_family="MER102"
repeat_region 39908..39928
/rpt_family="(CAA)n"
repeat_region complement(40118..40158)
/rpt_family="MSTA"
repeat_region 40159..40441

```

Query Match 17.9%; Score 19; DB 11; Length 220218;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 ACATTAATTAATTAATCA 97  
 |||||  
 DB 102394 AGATTAATTAATTAATCA 102412

```

RESULT 9
LOCUS HS135L22 176564 bp DNA HTG 09-MAY-1999
DEFINITION Homo sapiens chromosome 6 clone 135L22, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION AL031767
VERSION 94775594
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 176564)
AUTHORS Smalley,C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On May 11, 1999 this sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,

```

phage etc. Order of segments is not known; 800 n's separate  
 segments. Unfinished: dJ135L22 Contig\_ID: 00648 acc-AL031767  
 Length: 2037 bp Unfinished: dJ135L22 Contig\_ID: 01378  
 acc-AL031767 Length: 1241 bp Unfinished: dJ135L22 Contig\_ID:  
 02549 acc-AL031767 Length: 2664 bp Unfinished: dJ135L22  
 Contig\_ID: 03008 acc-AL031767 Length: 1039 bp Unfinished:  
 dJ135L22 Contig\_ID: 01030 acc-AL031767 Length: 1270 bp  
 Unfinished: dJ135L22 Contig\_ID: 00062 acc-AL031767 Length: 1104  
 bp Unfinished: dJ135L22 Contig\_ID: 02204 acc-AL031767  
 Length: 1306 bp Unfinished: dJ135L22 Contig\_ID: 01159 acc-AL031767  
 Length: 1253 bp Unfinished: dJ135L22 Contig\_ID: 02568  
 acc-AL031767 Length: 1541 bp Unfinished: dJ135L22 Contig\_ID:  
 01677 acc-AL031767 Length: 1024 bp Unfinished: dJ135L22  
 Contig\_ID: 01007 acc-AL031767 Length: 1560 bp Unfinished:  
 dJ135L22 Contig\_ID: 01326 acc-AL031767 Length: 1083 bp  
 Unfinished: dJ135L22 Contig\_ID: 01840 acc-AL031767 Length: 1452  
 bp Unfinished: dJ135L22 Contig\_ID: 00752 acc-AL031767 Length:  
 1018 bp Unfinished: dJ135L22 Contig\_ID: 02491 acc-AL031767  
 Length: 1170 bp Unfinished: dJ135L22 Contig\_ID: 03068  
 acc-AL031767 Length: 2473 bp Unfinished: dJ135L22 Contig\_ID:  
 02456 acc-AL031767 Length: 1126 bp Unfinished: dJ135L22  
 Contig\_ID: 03069 acc-AL031767 Length: 3213 bp Unfinished:  
 dJ135L22 Contig\_ID: 02097 acc-AL031767 Length: 1048 bp  
 Unfinished: dJ135L22 Contig\_ID: 00956 acc-AL031767 Length: 1139  
 bp Unfinished: dJ135L22 Contig\_ID: 01010 acc-AL031767 Length:  
 1494 bp Unfinished: dJ135L22 Contig\_ID: 00203 acc-AL031767  
 Length: 1457 bp Unfinished: dJ135L22 Contig\_ID: 00520  
 acc-AL031767 Length: 1051 bp Unfinished: dJ135L22 Contig\_ID:  
 00804 acc-AL031767 Length: 1145 bp Unfinished: dJ135L22  
 Contig\_ID: 02426 acc-AL031767 Length: 108892 bp Unfinished:  
 dJ135L22 Contig\_ID: 01175 acc-AL031767 Length: 12764 bp.  
 \* NOTE: This is a working draft sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

#### FEATURES

LOCATION/Qualifiers  
 1..176564  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="135L22"

BASE COUNT 44695 a 34144 c 32565 g 45093 t 20067 others  
 ORIGIN

Query Match 17.9%; Score 19; DB 34; Length 176564;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ACATTTTAAATTAATGTC 26  
 |||||  
 DB 147955 ACATTTTAAATTAATGTC 147937

```

RESULT 10
LOCUS HS501A4 79760 bp DNA HTG 12-MAY-1999
DEFINITION Homo sapiens chromosome X clone 501A4, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION Z98051
VERSION 94826497
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 79760)
AUTHORS Wray,P.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk

```

COMMENT

On May 13, 1999 this sequence version replaced g1:4007196.  
 IMPORTANT: This sequence is unfinished and does not necessarily  
 represent the correct sequence. Work on the sequence is in progress  
 and the release of this data is based on the understanding that the  
 sequence may change as work continues. The sequence may be  
 contaminated with foreign sequence from E.coli, yeast, vector,  
 phage etc. Order of segments is not known; 800 n's separate  
 segments. Unfinished: d3501A4 Contig\_ID: 00902 acc=298051  
 Length: 2580 bp unfinished: d3501A4 Contig\_ID: 01182 acc=298051  
 Length: 26213 bp unfinished: d3501A4 Contig\_ID: 00318 acc=298051  
 Length: 49367 bp.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

FEATURES

location/Qualifiers  
 1..79760  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /clone="501A4"

BASE COUNT 20651 a 17694 c 16902 g 22913 t 1600 others  
 ORIGIN

Query Match 17.9%; Score 19; DB 34; Length 79760;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OR 72 TTTTGTGACATTAATAAA 90  
 |||||  
 DB 74922 TTTTGTGACATTAATAAA 74904

RESULT 11

CERT12A7 39220 bp DNA INV 23-NOV-1998  
 LOCUS Caenorhabditis elegans cosmid T12A7, complete sequence.  
 DEFINITION 273911  
 ACCESSION 91359577  
 NID 273911.1 GI:1359577  
 VERSION Hrc: Lipase; Testis-specific protein TPX-1 like.  
 KEYWORDS Caenorhabditis elegans.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 39220)  
 Lemard N.  
 REFERENCE 1 Direct Submission  
 AUTHORS Submitted (02-JUN-1996) Louis, MO 63110, USA. E-mail:  
 TITLE jes@sanger.ac.uk or tw@nematode.wustl.edu  
 JOURNAL 2 (bases 1 to 39220)  
 REFERENCE 2 Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,  
 Bonfield, J., Burton, J., Connell, M., Copey, T., Cooper, J.,  
 Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,  
 Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,  
 Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N.,  
 Latreille, P., Lighthouse, J., Lloyd, C., McMurray, A., Mortimore, B.,  
 O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,  
 Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,  
 Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,  
 Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,  
 Wilkinson-Sproat, J. and Woldman, P.  
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans

TITLE Nature 368 (6466), 32-38 (1994)

JOURNAL MEDLINE 94150718  
 COMMENT Coding sequences below are predicted from computer analysis, using  
 predictions from GeneFINDER (P. Green, U. Washington), and other  
 available information.  
 For a graphical representation of this sequence and its analysis  
 see: -  
 http://webace.sanger.ac.uk/cgi-

FEATURES

source

gene  
 CDS

bin/display7db-wormaccetclass=Sequence object-T12A7  
 Current sequence finishing criteria for the C. elegans genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 IMPORTANT: This sequence is NOT necessarily the entire insert of  
 the specified clone. It may be shorter because we only sequence  
 overlapping sections once, or longer because we arrange for a small  
 overlap between neighbouring submissions.  
 This sequence is the entire insert of clone T12A7. The true right  
 end of clone K01A6 is at 1384 in this sequence. The start of this  
 sequence (1..110) overlaps with the end of sequence Z68750.  
 The end of this sequence (39117..39220) overlaps with the start of  
 sequence Z92847.

gene  
 CDS

1..39220  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="IV"  
 /clone="T12A7"  
 774..7967  
 /gene="T12A7.1"  
 3018(774..847,1121..1357,1414..1566,1877..2188,2236..2317,  
 2363..2878,2926..2991,3249..3680,4775..4913,4960..5071,  
 5316..5670,5723..5841,5886..5942,6810..7152,7428..7661,  
 7707..7967)  
 /gene="T12A7.1"  
 /note="Predicted using GeneFINDER; similar to C2 domain (2  
 domains)"  
 /codon\_start=1  
 /protein\_id="CA98141.1"  
 /db\_xref="PID:e1349250"  
 /db\_xref="PID:g3879788"  
 /db\_xref="GI:3879788"  
 /db\_xref="SPTREMBL:022422"  
 /translation="MNQPSGLTDSRPNKTVRLTISASNLMDVFSKSDPICLIE  
 KTSGRATTDPTAVTWTMDRQTEGRRTIINDNPOKTELLPYDEEQQLRF  
 ELTDVDSPKVAGMSGHDVLRGEYLAQIVSYSTRRAHLDGHSIGAOKNKDKYSK  
 NGSITITRAEDERNEKIQPDVCEGKELDRKSDPYLNFKRFPDGSHTMLHQEV  
 KPTLNPRTSTVLISDQALGKMDRPIIECDHREKFKODDIIETAGTQDLTLEQ  
 AGGESQIPLINEKKKLGKAKYKSGMKINNAVLVEPTFLDIAGTQDLFVA  
 VFTASGNPRTGSLHYMATGIPNOYEILRSVLTICOHYNSFTIEAYFGAKLPN  
 HQVSAVFPDILNROSSVYGVINGVNAVNAHALONTLTGPNSEPIIVEAVAKAKM  
 MKTARQIILITIDGTSMDRTINTVIRASLPSITIIIGVGNDFEPMHGLSDSD  
 ALAQOBSRIAEKRIQVPMRDVLGNOSGGLVTDVIMELAREVLEVPDQGLTKM  
 RRGFTPKSRDSDPMDRSPDEYDIPNDNLSIQASSPTGVPTIQLHATAPIINSY  
 NRIWRRLSDLAEPFRFSNDLKIPKIQKSLRTAMENVLAERMDAKLLEFVPA  
 KSKRRALPKSELLRTFYSKLVKRRHROVGDHVDVEMLMATINKDREFESSRL  
 PNOKVQCPFLAIRNGTIOLEVEYIMNOITDRDEALIGLEMRMAYRARGDAISFLC  
 HNCRRNALATCTCINAAAFDAFOHYHDKGKNDENKQYRKKEEPLKHCCEPLR  
 SLVAKHEHFRLLCDAEFRNEDELFSLHILFEELITVYLLIITILPSIGISTBEFL  
 SDRKIDGRYOLFVWCGLISTTMMHSLYLRKFIAPDPARKTIVKAAVADLVNPF  
 YLTSVLCCTCAIGVNMEDQKFAKNRIIEKFLVPPNMDNDVYTTDENPMKLYMYVT  
 VVEIILISIVIVPAIAFHLFSKRNREKSEKAEARLQALIFOMTILHFFHVP  
 LVQETWAIIRRTNNVITFLIGLFTMALHGAACAGIIVVKKPRLATKNHLKCYCRQK  
 VGKILITIVSYLSAR"  
 complement(8085..10201)  
 /gene="T12A7.3"  
 complement(3018(8085..8181,8459..8574,8627..8945,  
 9945..10201))  
 /note="Predicted using GeneFINDER; Similarity to Human  
 testis-specific protein TPX-1 (SW:TPX1\_HUMAN)"  
 /codon\_start=1  
 /protein\_id="CA98140.1"  
 /db\_xref="PID:e1349249"  
 /db\_xref="PID:g3879787"  
 /db\_xref="GI:3879787"  
 /db\_xref="SPTREMBL:022421"  
 /translation="MLIFLISVFIENVNGCGVCLINPPTAPRNRYPDVPRPD  
 PNQIPIDAOKCSDECDCEHEFMENVAFCQNCYVIRGSQLAFGNFTVTRHRR  
 AAGSNIKKFWMNATLERSAVSFAOKNPSQSFIPDIGENLFWMSTRPGDFNKYGPWA

ALSMIKEREFKFMDSNITLNDLFGSGVGHATQMWADTYQMGCAVSHKEKHKRGRP  
ITKICVCHWPKNGNLEPIYLEGPPCKESKCKDRKGLCTI  
gene complement(10372.14472)  
/gene="T12A7.2"  
complement(join(10372.10477,10529.10715,10758.10896,  
11897.12056,12565.12660,12717.13044,13882.14320,  
14365.14472))  
/gene="T12A7.2"  
/note="Predicted using GeneFinder; weak similarity to  
Mouse tenascin protein (TR:9220610)"  
/codon\_start=1  
/protein\_id="CA98142.1"  
/db\_xref="PID:e1349251"  
/db\_xref="PID:g3879789"  
/db\_xref="GI:3879789"  
/db\_xref="SPTREMBL:Q22423"  
/translation="MRSNSRLWVYGLMAYIIALLIYLSLPTPHLANNGMIVN  
DAHTEYAMKGOAYDMOETVKNHNSIGTSFPLREHRTITPASKOKFLNHISTVY  
QIKPEPNLGGPCRHGCVNTETDSCYCTKETSVCANDGCTFNNTLRCVC  
TSNNGEHGICIRCSNGVVKNTSGCECLGSPCTIOKICNGHFGKACACEGTGA  
CTIDGATPTIQCEVIRKRGVNSRLTQISFCITIGLCVGAHRRRAAMPVE  
ETWYRVHPNNRRPFCRHDMCGSWPDRALVAPRGSMSTNSRVHRLATPP  
SYTVDLNSAEOPTPEATRIENVALEISESEADDSVEPAKTLISDAE  
QEKHVKOLETLKLTFRIGGWEYAKRLRTYEMAMSEYDAVTEOHRKSEDAEA  
TTKQTYKNDCKRMKIGIKRKSINEQKSHAEVLAKEDEIATHSNADALPYTEVK  
AEIKSMKLPNK  
gene complement(23972.28860)  
/gene="T12A7.4"  
complement(join(23972.24058,24676.24752,25128.25296,  
25890.26020,26065.26219,26413.26495,271223.27321,  
28546.28860))  
/gene="T12A7.4"  
/note="Predicted using GeneFinder; similar to lipase"  
/codon\_start=1  
/protein\_id="CA98139.1"  
/db\_xref="PID:e1349248"  
/db\_xref="PID:g3879786"  
/db\_xref="GI:3879786"  
/db\_xref="SPTREMBL:Q22420"  
/translation="MILGERSLSYCYFEYFYVHTNTHIRALKFEVIFELFYRL  
LLVPSILISLVLSODIREGVPISANFPMAMDNGYKDYARLDLGTAGSGGSSMD  
MEYEDTSKDPYFFHNSDALTANNESTGWTTOYVLTGNGYTLGHXGSKGNTN  
TAYVRDHCVTYRLKRFVAMDYTGAKOINITSKGYTLAKRYLIGETINDDN  
SCNIGKPLGNRVILIGAGANFGIACAVGMGLVPTWCDNDLWGCETPYPPDN  
GQCAASPLSPCNGILYSEFLMDLNSRFRKAOHIFSMWSGDDILGCGNNVWGSPIS  
EIPFDSNRKIKYKMTKLTMTAADOFHITOLITP"  
join(34393.34506,34549.34688,34933.35033,35086.35178,  
35223.35366,35409.35545,35597.35715,35761.35867,  
36079.36285,36336.36671,36752.36958,37010.37161,  
37615.37742,37789.37916,37999.38078,38903.39033,  
292847.1:146.252,292847.1:972.1078,292847.1:2100.2364,  
292847.1:2409.2840)  
/gene="T12A7.5"  
/note="CDNA EST EMBL:D33084 comes from this gene; CDNA EST  
EMBL:D3335 comes from this gene; CDNA EST EMBL:D34203  
comes from this gene; CDNA EST EMBL:D55826 comes from this  
gene; CDNA EST EMBL:D36158 comes from this gene; CDNA EST  
EMBL:D36159 comes from this gene; CDNA EST EMBL:D64451  
comes from this gene; CDNA EST EMBL:D6446 comes from this  
gene; CDNA EST EMBL:D67354 comes from this gene; CDNA EST  
EMBL:D67441 comes from this gene; CDNA EST EMBL:D67491  
comes from this gene; CDNA EST EMBL:D68032 comes from this  
gene; CDNA EST yk18b7.3 comes from this gene; CDNA EST  
yk348b10.3 comes from this gene; CDNA EST yk348b10.5 comes  
from this gene"  
/codon\_start=1  
/protein\_id="CA98143.1"  
/db\_xref="PID:e1349252"  
/db\_xref="PID:g3879790"  
/db\_xref="GI:3879790"  
/db\_xref="SPTREMBL:O17904"  
/translation="MSFTFHFSILILIVISIEFSPCYKPIPLDLOPYNGSWP  
VIAPKPSASRIFLKNINSMISLILQNDKNTLSAEIHTYDQCHSLGSMTOHRKG

YMAEVNQTGVSNTYEKIFRSITGYSGEONEENNVLYGCLDNDNGCKITGPELV  
IMNSRHPQTLQILKSAQVLEDEACIDILHETADLTDECEDEVADDEQKIRHAKIDE  
DNEVDVCEKREYNITPANNINSLFEKREPRVITVAVIDASJSGDIAHISCRVETET  
AECMIRQDPCFNTNLTQTLGNSQIOISSLTNGSTGVTIDMGFVLMNGEDYISM  
HCGIYDDGDCGDSRVYVWSDHMDQPTINEIRTELQKVCVDDTDLIFLNTFECIT  
PSPTPNLRCSPPNGNSPITISMLQGWYIADNADPKMFLQSTVINMKRIQSSNL  
ELITYAOKESDGRICIGPONGVTKLTDGELGSRHSHNDLSMLPYLDKCIDKDLRWD  
NORAVLWCFERRSSNGCIGDVTFMRSRHSNDLSMLPYLDKCIDKDLRWD  
LHSOCGEISSSTILRDLITLTHYHVDITNVOEPCILNEIRGVADHSLERAG  
TWFLMSRNDKSMQTYAMVGRILFAVNTSTAVLRLFOGAAALGCEPDECTRYTIRER  
NDGSDYWELEHFEETSTNFMETFLFVRYHRYGVYSCLRNKPDCRCDKATVYS  
RHESIDAEELVLEKVAKSVCEETLHYAHABCELFENHRLKPPRCSIDSSNA  
EWEVLNQIEASTGRLLIOYVASSNGTSTFALRFQOHOKITTEBSGGLSPVSO  
IRFSLRRNLILYLPADGLVALIHPAIDHSDLLIREMGSCVCLAPALPEKONS  
OLIAFVCKLPNETLSTGOLNGDMWLYLSDPMFVLMNRCFTRDGDYSHTECFSES  
WVGCESTRCTINIHISNSPFIYFEEENAPIRTPAETISGRIQSVTDERLMLKE  
DHVGRAYSLMLRKTBEGMTEKVRQDAESFCVWPIESQIRHTALC"

Query Match 17.9% Score 19; DB 36; Length 39220;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 AGATAAATAAATAATCA 97  
DB 30907 AGATAAATAAATAATCA 30925

RESULT 12  
AC004722/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
COMMENT

AC004722 76619 bp DNA INV 29-AUG-1998  
Drosophila melanogaster DNA sequence (p1 DS02760 (D286)), complete  
sequence.  
AC004722  
93492861  
AC004722.1 GI:3492861  
HNC  
Drosophila melanogaster (subclones in sac from p1 clone DS02760  
(D286)) DNA.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 76619)  
Celinker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,  
Blazes,R.R., Harris,N.L., Agbayani,A., Accatna,T.T., Baxter,E.,  
Svirskas,R.R., Chavez,C., Chew,M., Doyle,L., Farfan,D.E.,  
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacled,U.M., Park,S.,  
Pfeiffer,B., Punch,E., Snit,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,  
Yee,A., Zhang,R., Zieran,L.L. and Rubin,G.M.  
Sequencing of Drosophila chromosome 2L, region 28C2-28C4  
Unpublished (1997)  
2 (bases 1 to 76619)  
Celinker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,  
Agbayani,A., Accatna,T.T., Baxter,E., Blazes,R.G., Chavez,C.,  
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,R.A.,  
Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,  
Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,  
Moshrefi,M., Nixon,K., Pacled,U.M., Park,S., Pfeiffer,B., Punch,E.,  
Snit,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,  
Zieran,L.L. and Kimmel,B.E.  
Direct Submission  
Submitted (22-MAY-1998) Drosophila genome center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Aug 29, 1998 this sequence version replaced g1:3419793.  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://fruitfly.berkeley.edu/sequence/>) or send

email to bddg@fruitfly.berkeley.edu.  
Library location: 72-29.

FEATURES  
SOURCE  
1. 76619  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="28C2-28C4"  
/clone="p1 DS05197 (D288)"  
BASE COUNT 22692 a 16312 c 16523 g 21092 t

ORIGIN  
BASE COUNT 22692 a 16312 c 16523 g 21092 t

Query Match 17.9%; Score 19; DB 37; Length 76619;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 ATAAATAAATATATCATT 99  
|||||  
Db 16429 ATAAATAAATATATCATT 16411

RESULT 13  
AC004724 47202 bp DNA INV 29-JUL-1998  
LOCUS Drosophila melanogaster DNA sequence (p1 DS05197 (D288)), complete  
DEFINITION  
AC004724 93347814  
NID AC004724.1 GI:3347814  
VERSION  
KEYWORDS  
SOURCE HTG.  
Drosophila melanogaster (Subclones in tet from p1 clone DS05197 (D288)) DNA.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS 1 (bases 1 to 47202)  
Celisner, S.E., George, R.A., Galle, R.F., Hoskins, R.A.,  
Svirskas, R.R., Harris, N.L., Aghayani, A., Arcaina, T.T., Baxter, E.,  
Blazek, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E.,  
Flanagan, J., Houston, K.A., Hummasti, S.R., Karris, K., Kearney, L.,  
Kim, S.H., Lee, B., Lomolan, M.A., Mak, J., Mazda, P., Mok, M.S.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S.,  
Pfeiffer, B., Punch, E., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R.,  
Yee, A., Zhang, R., Zieran, L.L., and Kimmel, B.  
Sequencing of Drosophila chromosome 2L, region 28C1-28C2  
Unpublished (1997)  
2 (bases 1 to 47202)  
Aghayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G., Chavez, C.,  
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,  
Hummasti, S.R., Karris, K., Kearney, L., Kim, S.H., Lee, B.,  
Lomolan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R.,  
Moshrefi, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B.,  
Punch, E., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A.,  
Zieran, L.L., and Kimmel, B.E.  
Direct Submission  
Submitted (22-MAY-1998) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Jul 29, 1998 this sequence version replaced g1:3287403.

TITLE  
JOURNAL  
COMMENT  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

FEATURES  
SOURCE  
1. 47202  
/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="28C1-28C2"  
/clone="p1 DS05197 (D288)"  
BASE COUNT 12918 a 9887 c 9989 g 14408 t

Query Match 17.9%; Score 19; DB 37; Length 47202;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 ATAAATAAATATATCATT 99  
|||||  
Db 25919 ATAAATAAATATATCATT 25937

RESULT 14  
AF011388 2307 bp DNA BCT 04-AUG-1997  
LOCUS Brevibacillus brevis cyclodextrin glycosyltransferase gene,  
DEFINITION  
AF011388 complete cds.  
ACCESSION  
AF011388 92293516  
NID AF011388.1 GI:2293516  
VERSION  
KEYWORDS  
SOURCE Brevibacillus brevis.  
ORGANISM Brevibacillus brevis  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Brevibacillus.

REFERENCE  
AUTHORS 1 (bases 1 to 2307)  
Kim, M.H., Sohn, C.B., and Oh, T.K.  
Cyclodextrin glycosyltransferase of Bacillus brevis CD162 and its  
structural gene  
Unpublished  
2 (bases 1 to 2307)  
Kim, M.H., Sohn, C.B., and Oh, T.K.  
Direct Submission  
Submitted (27-JUN-1997) Food and Nutrition, Chungnam National  
University, Yuseong, Taejeon 301-764, Korea

FEATURES  
SOURCE  
CDs  
1. 2307  
/organism="Brevibacillus brevis"  
/strain="CD162"  
/db\_xref="taxon:1393"  
1. 2079  
/EC\_number="2.4.1.19"  
/codon\_start=1  
/transl\_table=11  
/product="cyclodextrin glycosyltransferase"  
/protein\_id="AAB65420.1"  
/db\_xref="GI:2293517"

/db\_xref="GI:2293517"  
/translation="MTPSFIISSEFLPTVEASVTNKVSKDYIOIVTDRSDG  
HNPNGSGAIFSONGCDLKKYCGDQGIINKNDGYLIDGLITLMTISQVENVYAL  
HPGRTSYHAYNARVDYKTPNGSGDQDRLYTPNGKGIIMDFPNHSLLET  
NPVTEGALYNNAGLIGNYNDNRNKLPHNGGTDSSYEDSIVRLXDLADYLNK  
VDDOYKESIKLWLIKIDGIRVDAYKHAEGQGTSMSTIYKVPTEGEGFLDGE  
VDDOYKESIKLWLIKIDGIRVDAYKHAEGQGTSMSTIYKVPTEGEGFLDGE  
VDDOYKESIKLWLIKIDGIRVDAYKHAEGQGTSMSTIYKVPTEGEGFLDGE  
FINDHMSRSMVFNQDIALAVLLTSGVPTIYGEQVLTGNDPNKPKPT  
DRSNYSKIKTSKLSLRNSALGNTTERRWINSQVYIERFGNSVLTAVNSNR  
NQTISNNTSLPGNTDELQDNTITVANSAGNSPOLQANSVAWQVTKSTG  
PLIGHVGPATGKGTGYTVSGEGGKGSVLEGSTSAETVSNTEIGVNPYTAG  
HYNSVYVATNATNSPAREKEVYLSGNQSVRFVNNATNSGNTVYIGNVSELGMD  
PKNAIGPMQVQWKKPYTWYDLSVPGAKNLEKRIKKDHNGVNTWQSNNTYTPSA  
TGDTVTSNM"

BASE COUNT 738 a 453 c 466 g 650 t  
ORIGIN  
Query Match 17.0%; Score 18; DB 1; Length 2307;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:23:06 ; Search time 3059.95 Seconds  
(without alignments)  
68.331 Million cell updates/sec

Title: US-09-049-696-17

Perfect score: 106  
Sequence: 1 GGCAATCATTCTTTAAAT.....AAATTAATCATTCTCTT 106

Scoring table: OLIGO\_NWC

Searched: 2546578 seqs, 986266752 residues

Database : EST.\*

1: em\_est1:.\*  
2: em\_est2:.\*  
3: em\_est3:.\*  
4: em\_est4:.\*  
5: em\_est5:.\*  
6: em\_est6:.\*  
7: em\_est7:.\*  
8: em\_est8:.\*  
9: em\_est9:.\*  
10: em\_est10:.\*  
11: em\_est11:.\*  
12: em\_est12:.\*  
13: em\_est13:.\*  
14: em\_est14:.\*  
15: em\_est15:.\*  
16: em\_est16:.\*  
17: em\_est17:.\*  
18: em\_est18:.\*  
19: em\_est19:.\*  
20: em\_est20:.\*  
21: em\_est21:.\*  
22: em\_est22:.\*  
23: em\_est23:.\*  
24: em\_est24:.\*  
25: em\_est25:.\*  
26: em\_est26:.\*  
27: em\_est27:.\*  
28: em\_est28:.\*  
29: em\_est29:.\*  
30: em\_est30:.\*  
31: em\_est31:.\*  
32: em\_est32:.\*  
33: em\_est33:.\*  
34: em\_est34:.\*  
35: em\_est35:.\*  
36: em\_est36:.\*  
37: em\_est37:.\*  
38: em\_est38:.\*  
39: em\_est39:.\*  
40: em\_est40:.\*  
41: em\_est41:.\*  
42: em\_est42:.\*  
43: em\_est43:.\*  
44: em\_est44:.\*  
45: em\_est45:.\*  
46: em\_est46:.\*  
47: em\_est47:.\*  
48: em\_est48:.\*  
49: em\_est49:.\*  
50: em\_est50:.\*  
51: em\_est51:.\*  
52: em\_est52:.\*  
53: em\_est53:.\*

54: em\_est22:.\*  
55: em\_est23:.\*  
56: em\_est24:.\*  
57: em\_est25:.\*  
58: em\_est26:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	106	100.0	490	34	AA508854	AA508854 n122h10.s
C 2	106	100.0	716	49	A1660234	A1660234 w668902.x
C 3	106	100.0	734	49	A1660957	A1660957 w120d08.x
C 4	106	100.0	392	50	A1687981	A1687981 w377f06.x
C 5	98	92.5	161	48	A1582072	A1582072 at96a11.x
C 6	80	75.5	255	20	D25727	D25727 HMG504094
C 7	67	63.2	501	51	A1721275	A1721275 as82h08.x
C 8	65	61.3	335	51	A1721121	A1721121 as73d08.x
C 9	58	54.7	378	31	AA296955	AA296955 EST112726
C 10	18	17.0	232	25	N98035	N98035 2093c3 czap
C 11	18	17.0	544	41	A1041534	A1041534 ov82b07.x
C 12	18	17.0	528	41	A1042703	A1042703 uc76f07.x
C 13	18	17.0	528	41	A1042705	A1042705 uc76f09.x
C 14	18	17.0	509	44	A1288769	A1288769 qm06e04.x
C 15	18	17.0	421	50	A1692964	A1692964 w689h04.x
C 16	17	16.0	239	20	T37962	T37962 EST103279.S
C 17	17	16.0	380	20	T50928	T50928 yb88e01.r1
C 18	17	16.0	457	21	T86130	T86130 yd4c05.s1
C 19	17	16.0	198	24	D18949	D18949 M0SGS01139
C 20	17	16.0	311	24	H85858	H85858 y595f07.s1
C 21	17	16.0	478	24	N23257	N23257 y47e06.s1
C 22	17	16.0	445	25	N43088	N43088 5W3ICA1096S
C 23	17	16.0	480	27	AA001410	AA001410 z645d03.S
C 24	17	16.0	328	28	AA063691	AA063691 T3312 MVA
C 25	17	16.0	367	28	AA083382	AA083382 zn09h07.s
C 26	17	16.0	321	28	AA087603	AA087603 mm88d10.r
C 27	17	16.0	673	28	AA108534	AA108534 mp34h07.r
C 28	17	16.0	296	29	AA174679	AA174679 ms76e06.r
C 29	17	16.0	345	30	AA232941	AA232941 z146e01.S
C 30	17	16.0	360	30	AA243246	AA243246 z126f01.r
C 31	17	16.0	297	31	AA320173	AA320173 EST22495
C 32	17	16.0	227	33	AA389164	AA389164 vb24h08.r
C 33	17	16.0	283	33	AA452118	AA452118 zx15c04.s
C 34	17	16.0	222	34	AA490416	AA490416 as51c05.r
C 35	17	16.0	336	34	AA518393	AA518393 v126g10.r
C 36	17	16.0	595	36	AA565333	AA565333 vv14g07.r
C 37	17	16.0	392	36	C81399	C81399 C81399 Mous
C 38	17	16.0	267	39	AA823951	AA823951 vq93c12.r
C 39	17	16.0	450	42	A1150254	A1150254 qf34f07.x
C 40	17	16.0	587	42	AU023472	AU023472 AU023472
C 41	17	16.0	590	43	A1103282	A1103282 EST212571
C 42	17	16.0	381	43	A1191909	A1191909 qd63h03.x
C 43	17	16.0	846	43	A1194989	A1194989 u159c01.x
C 44	17	16.0	482	43	A1231519	A1231519 EST228207
C 45	17	16.0	405	44	A1246133	A1246133 qu38c07.x

#### ALIGNMENTS

RESULT 1  
LOCUS AA508854/c 490 bp mRNA  
DEFINITION n122h10.s1 NCI\_CGAP\_C04 Homo sapiens cDNA clone IMAGE:968803 3',  
ACCESSION AA508854  
NTD g2246357  
VERSION AA508854.1 GI:2246357

```

KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
              Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 490)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
              Unpublished (1997)
JOURNAL      On Sep 12, 1996 this sequence version replaced gi:1397407.
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LML at:
www-bio.lml.gov/bdnp/image/image.html

Insert Length: 856      Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 404.
Location/Qualifiers
    1..490
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone IMAGE:968803"
    /clone_1lb="NCI-CGAP_Co4"
    /sex="pooled"
    /tissue="type="colon"
    /lab_host="DH10B"
    /note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker: 1st strand cDNA was prepared from pooled colon
tumor tissue, and was then primed with a Not I - oligo(dt)
primer. Double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pRTT3 vector.
This library is not normalized. Library constructed by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT    155 a      82 c      89 g      164 t
ORIGIN
Query Match          100.0%; Score 106; DB 34; Length 490;
Best Local Similarity 100.0%; Pred. No. 2.1e-38;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GGCATTACATTTTAAATAATATGTCGAGTCGATGAGACCTGAGTCGATAGCC 60
        |||||||
DB      317 GGCATTACATTTTAAATAATATGTCGAGTCGATGAGACCTGAGTCGATAGCC 258

QY      61 TAGGCGCTGAATTTTGTCTCAGATTAATAATAATCATTCATCCTT 106
        |||||||
DB      257 TAGGCGCTGAATTTTGTCTCAGATTAATAATAATCATTCATCCTT 212

RESULT      2
LOCUS      AI660234/c
DEFINITION  AI660234 716 bp mRNA
            web8902.x1 Soares Dieckkraete.colon NHCD Homo sapiens cDNA clone
            IMAGE:2246290 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, MRNNA
            sequence.
ACCESSION  AI660234
NID        AI660234
VERSION    AI660234.1 GI:4763804
KEYWORDS   EST.
SOURCE     human
ORGANISM   Homo sapiens

```

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutharia; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 716)
JOURNAL	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997) On Mar 10, 1998 this sequence version replaced gi:2949219.
FEATURES	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1350 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. Seq primer: -40UP from Glbco High quality sequence stop: 387.  Location/Qualifiers 1..716 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2346290" /clone_1lb="Soares-Dieckgraefe_colon_NHCD" /tissue_type="colon:mucosa from 3 patients with Crohn's disease" /lab_host="DH10B (phage-resistant)" /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - 01190(dt) primer [5' TGTTACCAATCTGAAGTCGAGCGGCCGCCTGTATTATTTTTTTT 3'] , tightly-stranded cDNA was ligated to Eco RI adaptors (Pharmacia) , digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@um.wustl.edu); colon:mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo.
BASE COUNT	220 a 123 c 140 g 233 t
ORIGIN	
Query Match	100.0%; Score 106; DB 49; Length 716;
Best Local Similarity	100.0%; Prod. No. 1.7e-38;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 GGCCATTCACATTTTAAAAATTATGTGGAAGTGATAGAGAATCTCAGCTGTCAATACC 60 
Db	314 GGCAATTCACATTTTAAAAATTATGTGGAAGTGGAAGAGAAGAACCTCACATTAAGCC 255
QY	61 TAGGGCTGAATTTTGTTCAGATAAATAAATAATCATTCATCTT 106 
Db	254 TAGGGCTGAATTTTGTTCAGATAAATAAATAATCATTCATCTT 209
RESULT 3	
AI660957/c	AI660957 734 bp mRNA EST 10-MAY-1999
LOCUS	wf20d08.x1 Soares-Dieckgraefe.colon_NHCD_Homo sapiens cDNA clone
DEFINITION	IMAGE:2351151 3' similar to IR:088826 088826 GOB-5 PROTEIN.; mRNA
ACCESSION	sequence.
NID	AI660957
VERSION	G4764540
KEYWORDS	AI660957.1 GI:4764540
SOURCE	EST.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
	Eutharia; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 734)
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).



JOURNAL  
COMMENT  
Tumor Gene Index  
Unpublished (1997)  
On Mar 20, 1998 this sequence version replaced gi:2980033.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Glibco  
High quality sequence stop: 477.

## FEATURES

source

Location/Qualifiers  
1..734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="X"  
/clone IMAGE:2351151"  
/clone.lib="Soares.Dieckgraefe.colon\_NHUC"  
/tissue\_type="colonic mucosa from 5 ulcerative colitis  
patients"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pTR73D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TCTTCCCACTCTGAGTGGAGCGCCGCTAGCTTTTCTTTTCTTTT 3']  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pTR73 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieckelm.wustl.edu); colonic mucosa represents a range of  
ulceration, fibrosis, and degeneration. Library  
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 200 a 143 c 159 g 228 t 4 others  
ORIGIN

Query Match 100.0%; Score 106; DB 49; Length 734;  
Best local similarity 100.0%; Pred. No. 1.6e-38;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCAATTCACATTTTAAATAATATGTCGAGTCGATGAGAACTGCAGCTGTCAATGCC 60  
|||||  
Db 110 GGCAATTCACATTTTAAATAATATGTCGAGTCGATGAGAACTGCAGCTGTCAATGCC 51  
QY 61 TAGGCTGCAATTTTGTGCAGATAAATAATCAATTCATCCTT 106  
|||||  
Db 50 TAGGCTGCAATTTTGTGCAGATAAATAATCAATTCATCCTT 5

RESULT 4  
LOCUS A1687981 392 bp mRNA EST 27-MAY-1999  
DEFINITION wa77f06.x1 Soares.NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2802211 3' similar to TR:088826 O88826 GOB-5 PROTEIN.; mRNA  
sequence.  
ACCESSION A1687981  
NID 94899275  
VERSION A1687981.1 GI:4899275  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 392)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2947973.  
Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Glibco.

## FEATURES

source

Location/Qualifiers  
1..392  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone IMAGE:2302211"  
/clone.lib="Soares.NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pTR73D-Pac (Pharmacia) with  
a modified polylinker; Site.1: Not I; Site.2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCI CGAP GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo.  
BASE COUNT 125 a 69 c 67 g 131 t  
ORIGIN

Query Match 100.0%; Score 106; DB 50; Length 392;  
Best local similarity 100.0%; Pred. No. 2.4e-38;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCAATTCACATTTTAAATAATATGTCGAGTCGATGAGAACTGCAGCTGTCAATGCC 60  
|||||  
Db 281 GGCAATTCACATTTTAAATAATATGTCGAGTCGATGAGAACTGCAGCTGTCAATGCC 222  
QY 61 TAGGCTGCAATTTTGTGCAGATAAATAATCAATTCATCCTT 106  
|||||  
Db 221 TAGGCTGCAATTTTGTGCAGATAAATAATCAATTCATCCTT 176

RESULT 5  
LOCUS A1582072 161 bp mRNA EST 06-APR-1999  
DEFINITION ar96a11.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
IMAGE:2173436 3' mRNA sequence.  
ACCESSION A1582072  
NID 94567969  
VERSION A1582072.1 GI:4567969  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 161)  
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Getseil,G., Jost,S.,  
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Mar 19, 1997 this sequence version replaced gi:1900677.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Glibco  
High quality sequence stop: 115.

QY	1	GGCATTGCACATTTTAAAAATATATGTGGAAGTGAGTAGAGAACTGCACCTGTCAATATACC	60
DB	133	GGCATTGCACATTTTAAAAATATATGTGGAAGTGAGTAGAGAACTGCACCTGTCAATATACC	192
QY	61	TAGGCGCTGAATTTTGTGCAG	80
DB	193	TAGGCGCTGAATTTTGTGCAG	212
RESULT	7		
LOCUS	AI721275/c		
DEFINITION	ase2h08.x1 Barstead colon HPLRB7 Homo sapiens EST	10-JUN-1999	
ACCESSION	IMAGE:2333263.3	similar to SW:ECOLC_BOVIN F54281 EPITHELIAL CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.	
NID	AI721275		
VERSION	95038531		
KEYWORDS	AI721275.1	GI:5038531	
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 501)		
JOURNAL	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S., Kitzman, J., Moore, B., Schellenberg, K., Stepec, M., Tan, F., Thaising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.		
COMMENT	WashU-NCI human EST Project		
	Unpublished (1997)		
	On Jun 22, 1998 this sequence version replaced gi:3246918.		
FEATURES			
source	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: east@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Possible reversed clone: similarity on wrong strand Seq primer: -40UP from Gibco High quality sequence stop: 394. Location/Qualifiers 1..501 /organism="Homo sapiens" /db_xref="taxon:9606" /map="Xp11.2" /clone="IMAGE:2335263" /clone_id="Barstead colon HPLRB7" /sex="male" /dev_stage="adult, age 25" /lab_host="DH10B (phage resistant)" /note="Organ: colon; Vector: pTV3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTAGCAATCGTAGGAGGAGGCGGCCGCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [5' AATTCACCTAGTAT 3' and 5' ATTCAATCG 3'], digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV3 vector. Library constructed by Bob Barstead."		
BASE COUNT	158 a	83 c	106 g
ORIGIN			154 t
Query Match	63.2%;	Score 67;	DB 51; Length 501;
Best Local Similarity	100.0%;	Pred. No. 5, 8e-21;	
Matches	67;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	GGCATTGCACATTTTAAAAATATATGTGGAAGTGAGTAGAGAACTGCACCTGTCAATATACC	60
DB	123	GGCATTGCACATTTTAAAAATATATGTGGAAGTGAGTAGAGAACTGCACCTGTCAATATACC	64

OY	61	TAGGACT	67
Db	63	TAGGCT	57
RESULT	8		
A1721121/c			
LOCUS			
DEFINITION	A1721121	335 bp	mRNA
	as73d08.x1	Bartstead colon HPLRB7	Homo sapiens CDNA clone
	IMAGE:2334351	3'	similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA
sequence.			
ACCESSION	A1721121		
NID	95038377		
VERSION	A1721121.1	GI:5038377	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 335)		
	Hillier,L., Allen,M., Bowles,L., Dubnue,T., Geisel,G., Jost,S.,		
	Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Matra,M.,		
	Mattlin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,		
	Theisinger,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.		
TITLE	WashU-NCI human EST Project		
JOURNAL	Unpublished (1997)		
COMMENT	On Jun 22, 1998 this sequence version replaced gi:3246762.		
	Contact: Wilson RK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@wustl.edu		
	This clone is available royalty-free through LINL; contact the		
	IMAGE Consortium (info@image.lnl.gov) for further information.		
	Sequence reversed clone: similarity on wrong strand		
	Seq primer: -40UP from Glbo		
	High quality sequence stop: 312.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..335		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2334351"		
	/clone_lib="Bartstead colon HPLRB7"		
	/sex="male"		
	/dev_stage="adult, age 25"		
	/lab_host="DH10B (phage resistant)"		
	/note="Organ: colon; Vector: pRTD-Pac (Pharmacia) with a		
	modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st		
	strand cDNA was primed with a Not I - oligo(dT) primer [5		
	TGTATCACTGAAGTAGGAGCGGCCCCCTTTTTTTTTTTTTTTTTTTTTT		
	3'] double-stranded cDNA was ligated to Eco RI adaptors		
	[5' ATTCACTAATAAT 3' and 5' ATTCACTAGC 3'], digested		
	with Not I and cloned into the Not I and Eco RI sites of		
	the modified pRTD vector. Library constructed by Bob		
	Bartstead."		
BASE COUNT	98 a	57 c	71 g
ORIGIN			109 t
Query Match	61.3%; Score 65; DB 51; Length 335;		
Best Local Similarity	100.0%; Pred. No. 5.7e-20;		
Matches	65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1	GGCATTCAATTTAAATAATATGTGGGAAGTGATAGAGAACAATGCAATAGCC	60
Db	104	GGCATTCAATTTAAATAATATGTGGGAAGTGATAGAGAACAATGCAATAGCC	45
OY	61	TAGG	65
Db	44	TAGG	40

RESULT	10
989035	
LOCUS	AA296955
DEFINITION	EST112726 Colon I Homo sapiens cDNA 5' end, mRNA sequence.
NID	AA296955
VERSION	AA296955.1 GI:1949515
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 378) Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Feldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cocton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,D.L., Geoghegan,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelleys,J.M., Kelley,J.C., Liu,L.-I., Marrairos,S.M., Merrick,J.M., Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utecherbeck,T.R., Weldon,J.P., Li,Y., Benadark,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dinke,D., Feng,D.-F., Ferlie,A., Fischer,C., Hastings,G.A., He,M.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Melsner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Pannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence Nature 377 (6547 Suppl), 3-174 (1995) 96026280 On Sep 12, 1996 this sequence version replaced gi:1288227. Other_ESTs: THC167738 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlav@tigr.org For clone availability, additional sequence and expression Information related to this EST, please check the TIGR Human Gene Index ( <a href="http://www.tigr.org/tcdb/hgi1.html">http://www.tigr.org/tcdb/hgi1.html</a> ) Seq primer: M13 Reverse.
FEATURES	Location/Qualifiers
SOURCE	1..378 /organism="Homo sapiens" /db_xref="ATCC (Inhost):127834" /db_xref="taxon:9606" /clone_lib="Colon I" /dev_stage="adult" /note="Organ: colon; Ecoli: Site_2: XhoI"
BASE COUNT	112 a 79 c 64 g 119 t 4 others
ORIGIN	
Query Match	54.7%; Score 58; DB 31; Length 378;
Best Local Similarity	100.0%; Pred. No. 7.2e-17;
Matches	58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 GGCACTCACATTAAAATAATGATGTGGGAAGTGGATGAGAAGAACCTGCACATAG 58
DB	265 GGCACTCACATTAAAATAATGATGTGGGAAGTGGATGAGAAGAACCTGCACATAG 322

LOCUS N98035 232 bp mRNA EST 18-NOV-1996  
 DEFINITION 2093C3 czappPd2.1, Debopam Chakrabarti Plasmodium falciparum CDNA  
 clone PF2093C, mRNA sequence.  
 ACCESSION N98035  
 NID 91675069  
 VERSION N98035.1 GI:1675069  
 KEYWORDS EST.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 232)  
 AUTHORS Dame J.B., Arnot D.E., Bourke P., Chakrabarti D., Christodoulou Z.,  
 Coppel R., Cowman A., Craig A., Fischer K., Foster J., Goodman N.,  
 Hildeberg K., Holder A.A., Holt D., Kemp D., Lanzer M., Lim A.,  
 Newbold C., Ravetch J.V., Reddy G.R., Rubio J., Schuster S.M.,  
 Su X.-Z., Thompson J.K., Vital F., Wellens T.E. and Werner E.  
 TITLE Current status of the Plasmodium falciparum genome project  
 JOURNAL Mol. Biochem. Parasitol. 79, 1-12 (1996)  
 MEDLINE 97001675  
 COMMENT On May 9, 1995 this sequence version replaced gi:802415.  
 CONTACT: Debopam Chakrabarti  
 Department of Molecular Biology and Microbiology  
 University of Central Florida  
 Orlando, FL 32816-2360  
 Tel: 407 384 2061  
 Fax: 407 384 3095  
 Email: dechakepegasus.cc.ucf.edu  
 Seq primer: T3.  
 FEATURES  
 source  
 location/Qualifiers  
 1..232  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /strain="Dd2"  
 /map="12p"  
 /clone="PF2093C"  
 /clone\_lib="czappPd2.1, Debopam Chakrabarti"  
 /lab\_host="E. coli XL-1 blue"  
 /note="Vector: Lambda ZAP II; Site\_1: EcoR I; Site\_2: Xho  
 I; POLYA+ RNA, from asynchronous blood stage parasites of  
 the Dd2 isolate cultured in vitro, was reverse transcribed  
 using an oligo dt-Xho I primer. Second strand was  
 prepared using RNase H and DNA polymerase I. EcoR I  
 adapters were ligated to the cDNA, and it was digested  
 with Xho I. Prepared fragments were ligated into EcoR I +  
 Xho I digested lambda ZAP II vector." 3 others  
 BASE COUNT 108 a 23 c 22 g 76 t 3 others  
 ORIGIN  
 Query Match 17.0%; Score 18; DB 25; Length 232;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 8 ACATTTTAAATATATGT 25  
 Db 158 ACATTTTAAATATATGT 175  
 RESULT 11  
 LOCUS A1041534 544 bp mRNA EST 27-AUG-1998  
 DEFINITION 0V82B07.x1 Soares\_testis\_NHT Homo sapiens CDNA clone IMAGE:1643797  
 3', mRNA sequence.  
 ACCESSION A1041534  
 NID 93280728  
 VERSION A1041534.1 GI:3280728  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 544)

AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Jan 19, 1998 this sequence version replaced gi:2152781.  
 CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 clone distribution: NCI-CCAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
 www-bio.linn.gov/bdrp/image/image.html  
 Insert Length: 1154 Std Error: 0.00  
 Seq primer: -40m13 fwd. E7 from Amersham  
 High quality sequence stop: 445.  
 FEATURES  
 source  
 location/Qualifiers  
 1..544  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1643797"  
 /clone\_lib="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pTR73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech  
 Laboratories, Inc., and primed with a Not I - oligo(dt)  
 primer [5',  
 TGTACCAATCTGAGTGGAGCGCCGCCCAATTTTTTTTTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pTR73 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo." 153 t  
 BASE COUNT 181 a 73 c 137 g 153 t  
 ORIGIN  
 Query Match 17.0%; Score 18; DB 41; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 6 TCACATTTTAAATATAT 23  
 Db 36 TCACATTTTAAATATAT 19  
 RESULT 12  
 LOCUS A1042703 528 bp mRNA EST 01-JUL-1998  
 DEFINITION uc76f07.x1 Sugano mouse liver m1a Mus musculus CDNA clone  
 IMAGE:1431589 3', mRNA sequence.  
 ACCESSION A1042703  
 NID 93285963  
 VERSION A1042703.1 GI:3285963  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 528)  
 AUTHORS Maier M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,  
 Geisler S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,  
 Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,  
 Thelning B., Wylie T., Lennon G., Soares B., Wilson R. and  
 Waterston R.  
 TITLE The MASHU-HMT Mouse EST Project  
 JOURNAL Unpublished (1996)

COMMENT On Jan 19, 1998 this sequence version replaced g1:2153097.

Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:915657

Seq primer: custom primer used  
High quality sequence stop: 453.  
Location/Qualifiers

#### FEATURES

##### source

1..528  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/map="829E06:3:3p14.3-3p21.1:17q21"  
/clone="IMAGE:1431589"  
/clone\_1lb="Sugano mouse liver mlia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: liver; Vector: PME18S-FL3; Site\_1: DraIII  
(CAGCTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA  
was primed with an oligo(dT) primer  
[ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was  
ligated to a DraIII adaptor (TGTGGCCCTACTGG), digested  
and cloned into distinct DraIII sites of the PME18S-FL3  
vector (5' site CAGCTGTG, 3' site CACCATGTG). XhoI should  
be used to isolate the cDNA insert. Size selection was  
performed to exclude fragments <1.5kb. Library  
constructed by Dr. Sumio Sugano (University of Tokyo  
Institute of Medical Science). Custom primers for  
sequencing: 5' end primer CTTCGTCTCTMAAGCTGCG and 3' end  
primer CGACCTGACGCTGACGACA."

BASE COUNT  
ORIGIN

167 a 71 c 126 g 164 t

Query Match 17.0%; Score 18; DB 41; Length 528;  
Best Local Similarity 100.0%; Pred. No. 47;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TCACATTTTAAATAT 23  
|||||  
DB 48 TCACATTTTAAATAT 31

RESULT 13  
AI042705/c 528 bp mRNA EST 01-JUL-1998  
LOCUS uc76f09.x1 Sugano mouse liver mlia Mus musculus cDNA clone  
DEFINITION IMAGE:1431593 3', mRNA sequence.  
ACCESSION AI042705  
NID 93285965  
VERSION AI042705.1 GI:3285965  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 528)  
AUTHORS Marie,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 19, 1998 this sequence version replaced g1:2152761.

Contact: Maria M/Mouse EST Project

WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:915651

Seq primer: custom primer used  
High quality sequence stop: 453.  
Location/Qualifiers

#### FEATURES

##### source

1..528  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/map="17q21"  
/clone="IMAGE:1431593"  
/clone\_1lb="Sugano mouse liver mlia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: liver; Vector: PME18S-FL3; Site\_1: DraIII  
(CAGCTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA  
was primed with an oligo(dT) primer  
[ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was  
ligated to a DraIII adaptor (TGTGGCCCTACTGG), digested  
and cloned into distinct DraIII sites of the PME18S-FL3  
vector (5' site CAGCTGTG, 3' site CACCATGTG). XhoI should  
be used to isolate the cDNA insert. Size selection was  
performed to exclude fragments <1.5kb. Library  
constructed by Dr. Sumio Sugano (University of Tokyo  
Institute of Medical Science). Custom primers for  
sequencing: 5' end primer CTTCGTCTCTMAAGCTGCG and 3' end  
primer CGACCTGACGCTGACGACA."

BASE COUNT  
ORIGIN

167 a 71 c 126 g 164 t

Query Match 17.0%; Score 18; DB 41; Length 528;  
Best Local Similarity 100.0%; Pred. No. 47;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TCACATTTTAAATAT 23  
|||||  
DB 48 TCACATTTTAAATAT 31

RESULT 14  
AI288769/c 509 bp mRNA EST 29-JAN-1999  
LOCUS gnm06e04.x1 NCI-CGAP\_lu5 Homo sapiens cDNA clone IMAGE:1881054 3',  
DEFINITION mRNA sequence.  
ACCESSION AI288769  
NID 9393272  
VERSION AI288769.1 GI:3932272  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 509)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)

COMMENT On Aug 21, 1998 this sequence version replaced.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov

Tissue Procurement: Christopher Moskalk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbrr/image/image.html](http://www.bio.llnl.gov/dbrr/image/image.html)

Insert Length: 553 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 426.  
 Location/Qualifiers  
 1. 509

FEATURES  
 source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1881054"  
 /clone\_1ib="NCI-CGAP\_Lu5"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 156 a 90 c 82 g 181 t

ORIGIN

Query Match 17.0%; Score 18; DB 44; Length 509;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 TCAGATTAATAAATAA 94  
 ||||||||||||||||  
 DB 338 TCAGATTAATAAATAA 321

RESULT 15  
 A1692964/C 421 bp mRNA EST 02-JUN-1999  
 LOCUS wd89h04.x1 NCI-CGAP\_Lu24 Homo sapiens CDNA clone IMAGE:2338807 3',  
 DEFINITION mRNA sequence.  
 ACCESSION A1692964  
 NID 94970304  
 VERSION A1692964.1 GI:4970304  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 421)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 JOURNAL On Jun 5, 1998 this sequence version replaced gi:3188051.  
 COMMENT

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
 Tissue Procurement: Christopher Moskalk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbrr/image/image.html](http://www.bio.llnl.gov/dbrr/image/image.html)

FEATURES  
 source  
 Seq primer: -40UP from Gibco.  
 Location/Qualifiers  
 1. 421

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2338807"  
 /clone\_1ib="NCI-CGAP\_Lu24"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP\_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneds 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 133 a 67 c 63 g 158 t

ORIGIN

Query Match 17.0%; Score 18; DB 50; Length 421;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 TCAGATTAATAAATAA 94  
 ||||||||||||||||  
 DB 347 TCAGATTAATAAATAA 330

Search completed: August 6, 1999, 09:23:08  
 Job time: 6259 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:16:00 ; Search time 3264.21 Seconds

(without alignments)  
103.275 Million cell updates/sec

Title: US-09-049-696-17

Perfect score: 106  
Sequence: 1 GGCATTCACATTTTAAAAAT.....AAATAATCATTCATCCTT 106

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.\*

1: gb\_dal:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_p11:\*  
8: gb\_p12:\*  
9: gb\_p13:\*  
10: gb\_p14:\*  
11: gb\_p15:\*  
12: gb\_p16:\*  
13: gb\_p17:\*  
14: gb\_p18:\*  
15: gb\_p19:\*  
16: gb\_p20:\*  
17: gb\_p21:\*  
18: gb\_p22:\*  
19: gb\_p23:\*  
20: gb\_p24:\*  
21: gb\_p25:\*  
22: gb\_p26:\*  
23: gb\_p27:\*  
24: gb\_p28:\*  
25: gb\_p29:\*  
26: gb\_p30:\*  
27: gb\_p31:\*  
28: gb\_p32:\*  
29: gb\_p33:\*  
30: gb\_p34:\*  
31: gb\_p35:\*  
32: gb\_p36:\*  
33: gb\_p37:\*  
34: gb\_p38:\*  
35: gb\_p39:\*  
36: gb\_p40:\*  
37: gb\_p41:\*  
38: gb\_p42:\*  
39: gb\_p43:\*  
40: gb\_p44:\*  
41: gb\_p45:\*  
42: gb\_p46:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	106	100.0	3311	11	AF039400	Homo sapi

2	106	100.0	35278	11	AF039401	AF039401 Homo sapi
3	106	100.0	2826	42	AF127036	AF127036 Homo sapi
4	76.8	72.5	878	5	195746	195746 Sequence 8
5	47.4	44.7	2937	12	AB017156	AB017156 Mus muscu
6	36.6	34.5	29324	1	BS16829KB	AJ225287 Bacillus
7	36.6	34.5	3215	1	BSCHEVE	Z29584 B. subtilis
8	36.6	34.5	3215	1	BSU05345	U05345 Bacillus su
9	36.6	34.5	208230	1	BSUB0008	Z29111 Bacillus su
10	34	32.1	12984	2	AE001122	AE001122 Borrelia
11	34	32.1	110000	34	CEY113B8_1	Continuation (2 of
12	34	32.1	224129	34	CEY7A9	Z295312 Caenorhabd
13	34	32.1	26821	36	CEY7A9C	Z29286 Caenorhabd
14	33.6	31.7	30567	36	CET28F4	Z72517 Caenorhabd
15	32.4	30.6	2508	1	STRSCRA	M22717 S. mutans en
16	31.8	30.0	103916	11	AC004225	AC004225 Homo sapi
17	31.8	30.0	175893	34	AC007004	AC007004 Homo sapi
18	31.8	30.0	175916	35	AC007747	AC007747 Homo sapi
19	31	29.2	256549	35	AC007564	AC007564 Homo sapi
20	30.8	29.1	14943	1	SPCPS14E	X85787 S. pneumonia
21	30.8	29.1	256172	35	AC005139	AC005139 Plasmodu
22	30.8	29.1	284972	35	AC005505	AC005505 Plasmodu
23	30.2	28.5	85214	8	ATAC005396	ATAC005396 Arabidops
24	30.2	28.5	77439	9	HS127E18	AL031575 Human DNA
25	30.2	28.5	166531	10	HSAC002123	AC002123 Human BAC
26	30.2	28.5	110000	34	CEY45F10_0	Z23245 Caenorhabd
27	30.2	28.5	191908	34	CEY51B9	Z28947 Caenorhabd
28	30.2	28.5	256014	34	CEY62E10	AL031580 Caenorhab
29	30.2	28.5	39271	36	CEY52G2	Z82269 Caenorhabd
30	30.2	28.5	28220	36	CEM176	Z78412 Caenorhabd
31	30.2	28.5	47042	42	AC007031	AC007031 Homo sapi
32	30	28.3	113762	11	AC005294	AC005294 Homo sapi
33	28.8	28.1	138708	36	CELC16C4	AF016657 Caenorhab
34	29.6	27.9	137740	1	D90900	D90900 Synchocyst
35	29.6	27.9	8212	7	ALR1BPT	X75653 A. longa pla
36	29.6	27.9	100000	9	AP000039	AP000039 Homo sapi
37	29.6	27.9	100000	9	AP000107	AP000107 Homo sapi
38	29.6	27.9	100000	9	AP000183	AP000183 Homo sapi
39	28.6	27.9	68863	9	AP000280	AP000280 Homo sapi
40	28.6	27.9	215609	11	AC002485	AC002485 Human PAC
41	29.6	27.9	102008	11	AC005228	AC005228 Homo sapi
42	29.4	27.7	116160	9	AB015752	AB015752 Homo sapi
43	29.4	27.7	122121	11	AC004543	AC004543 Homo sapi
44	29.2	27.5	3213	8	AF070480	AF070480 Emeritell
45	29.2	27.5	86618	10	HS593F22	AL030999 Human DNA

## ALIGNMENTS

RESULT 1  
AF039400 3311 bp mRNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hclcal) mRNA,  
DEFINITION complete cds.  
ACCESSION AF039400  
NTD 94009457  
VERSION AF039400.1 GI:4009457  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,R.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human hclcal, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission.

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers  
SOURCE

1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 106; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 2e-20;

Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCATTCATTTTAAATATATGCGAAGTGTAGAGAACTGCTCATACCC 60  
|||||  
Db 3034 GGCATTCATTTTAAATATATGCGAAGTGTAGAGAACTGCTCATACCC 3093  
QY 61 TAGGGCTGAATTTTGTGAGATAAATAATCAATCATCTT 106  
|||||  
Db 3094 TAGGGCTGAATTTTGTGAGATAAATAATCAATCATCTT 3139

RESULT 2  
AF039401  
LOCUS AF039401 35278 bp DNA PRI 15-DEC-1998  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.  
ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009458  
KEYWORDS  
SOURCE human.  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human CLCA1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)

MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers  
SOURCE

promoter  
gene  
repeat\_region  
TATA\_signal  
exon  
mRNA  
CDS  
exon  
exon

1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1. .1617  
/gene="hCLCA1"  
1. .33522  
/gene="hCLCA1"  
767. .1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
/gene="hCLCA1"  
/note="putative"  
1618. .1708  
/gene="hCLCA1"  
/number=1  
join(1618. .1708,1962. .2383,6667. .6807,6979. .7126,9696. .9801,  
15454. .15631,18591. .18809,19783. .20010,22253. .22427,  
24522. .24628,26644. .26859,27447. .27708,28766. .28936,  
31840. .32079,32919. .33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962. .2383  
/gene="hCLCA1"  
/number=2  
join(2222. .2383,6667. .6807,6979. .7126,9696. .9801,  
15454. .15631,18591. .18809,19783. .20010,22253. .22427,  
24522. .24628,26644. .26859,27447. .27708,28766. .28936,  
31840. .32079,32919. .33522)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPKSSVFTLLHLLEGALSLSLIQLNNNGYEGIVADPNVP  
EDETLIQIKDMVTQASLYLFEATGKRFYFNVAIILPETWTKADYVRKLETYNA  
DYLVAESTPGNDPEPTDOMNGCEGGERIHLPDFAKKLAEGYPOGKAFFEMAH  
LRMGVDEYNDEKFLNSGRIOAVRCSAGITGVYKCKCGSGCYTKRCFTKVTGL  
YKGCFFVLOSROTERKASIMFAOHVDSIVECTEONHNKAPKONOKNLSRWETI  
RDSDEKRTKTPMTOPNPFTSLQIGRTICVLKSGSMATGNRLNNOGQFL  
LQVTELSWGWMTFDSAAHVOSELIOINSQSDRTLAKRLPAASGGSITCSGLSA  
FTVIRKRYPTDGSSEIVLTDGEDNTISGCFNEVKSQSAIHTVALGSAQDELISK  
MTGGLQTVASDOVONNGLIDAFALSSGNGAVSORIOLESKGLTQNSQMNQVTVY  
DSYVGDITLFLITWTPQIILMDPSGKGGFVVDKNTKMAVLIQIPGIKQNTWY  
SIQASSOTLITVTSRASNAATLPITVTSKNTKDSKPSPLVYVANSVRAAGPILRA  
SVTALISVNGKVTLELDNGAGATATDDGVYSRTFTYDTNGRYSKVRPAIGGVN  
AARRRTIPOOSGALITPGWINDETIONMPPEPINKDDVQHKVCSRTSSGGSFVAS  
DVPNAPIDLPFPQGITDLNAETHGSLINLTWAPGDYDGHGTAKYIIRISTSLD  
LRDKFNESLQVNTALIPKEANSEVLFKRNITFENGDTDLFLAIOAVKVDLKEEI  
SNIAKVSLEFIPQTPETPSPDETSAPCPNHINSTIPGIHLIKIMKMWIGELQSLIA  
"  
6667. .6807  
/gene="hCLCA1"  
/number=3  
6979. .7126  
/gene="hCLCA1"  
/number=4  
9696. .9801  
/gene="hCLCA1"  
/number=5  
15454. .15631  
/gene="hCLCA1"  
/number=6



exon 18591..18809  
/gene="hCLCA1"  
/number=7  
exon 19783..20010  
/gene="hCLCA1"  
/number=8  
exon 22253..22427  
/gene="hCLCA1"  
/number=9  
exon 24522..24628  
/gene="hCLCA1"  
/number=10  
exon 26644..26859  
/gene="hCLCA1"  
/number=11  
exon 27447..27708  
/gene="hCLCA1"  
/number=12  
exon 28766..28936  
/gene="hCLCA1"  
/number=13  
exon 31840..32079  
/gene="hCLCA1"  
/number=14  
exon 32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN

Query Match 100.0%; Score 106; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 1.9e-20;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCATTCACATTTTAAATATATGGAAGTGGATAGAGACGCGCTGCAATAGCC 60  
DB 33248 GGCATTCACATTTTAAATATATGGAAGTGGATAGAGACGCGCTGCAATAGCC 33307  
QY 61 TAGGCGTGAATTTTGTACAGATAAATAAATCAATTCATCCTT 106  
DB 33308 TAGGCGTGAATTTTGTACAGATAAATAAATCAATTCATCCTT 33353

RESULT 3  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcicium-activated chloride channel protein 1 (CACCL1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2826)  
AUTHORS Agnel,M., Vernalt,T. and Culouscou,J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride  
CHANNEL Unpublished  
JOURNAL 2 (bases 1 to 2826)  
REFERENCE Agnel,M. and Culouscou,J.-M.  
AUTHORS Direct Submission  
TITLE Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
JOURNAL Carrières, Neuilly-Malmaison 92500, France  
FEATURES  
source  
1..2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1..2826  
gene

CDS  
/gene="CACCL1"  
5..2749  
/gene="CACCL1"  
/note="Dovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPKSSVFILILHLLEGALSNSLIQLNNGNVEGIVVAIDPNVP  
EDELILQIKDWTQASLYLFEATGRKRFKAVALLIPETWTKADYVPKLETKNA  
DYLVASSTPPGNDDEPTBOMGNGCKGEIHLTPPEIAGKTKLAIEGPOGARVHEHA  
LRMGVTFDEINDEKFTLSNGRIQAVRCSAGITGVNVKCKOGSCCTKCTKNTGL  
YKRGCEFLVQSQTTERASTIMEFAHSDIVIEFTEDHNEAPNKKONKLNSTVEVI  
RSEDEKFTTPTQPNPTFSLQIGRIYGLVLDKSGMAITGNLNLNCGQLFL  
LQTVELGSGVWGVTFDSAAHVQSELIQINSGLDSDPTLAKRLPAAASGGTSISGLISA  
FTVIRKKYPTDGESEIVLLTDEGNTISGCFNFKOSGAIHTVALGPSAOLLEELISA  
MTGGLQTVASDOVONNGILIDARGALISGNGAVSORSIOLESKGLTLONSOMNGFTYV  
DSTVGKDTLFLTWTTPQPOILLMPSSGCKGFTYVDKRTKAYLIQIPGIATVGTWKY  
SLQASQTLTLTLTVTSRASNATLPITVTSKTKNDISKFPPLVITANIKQASPIURA  
SVTALIESVNGKRTVLELDNGAGADATKDDVYSRYFTTYDINGRYSVKVAALGVN  
AARRVYIPQSSGALYIPGWIEENDEIQMNPPEINKDDVOHQVCFSTSSGSEFVAS  
DVNPAPIPLDFPFGQITDLKAEIHGSSILNLTWTPAGDDYDGHATKYYIIRISTILD  
LDRKNESLOVNTALIPKENSEVEFLKFNITFENGTDLEFLIAQAVDYLKSEI  
SNIAVSLFIPQTPETPSPDETSAPCNHININSTIPSHILIKIMKIKGELQLSIA

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 106; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 2e-20;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCATTCACATTTTAAATATATGGAAGTGGATAGAGACGCGCTGCAATAGCC 60  
DB 2687 GGCATTCACATTTTAAATATATGGAAGTGGATAGAGACGCGCTGCAATAGCC 2746  
QY 61 TAGGCGTGAATTTTGTACAGATAAATAAATCAATTCATCCTT 106  
DB 2747 TAGGCGTGAATTTTGTACAGATAAATAAATCAATTCATCCTT 2792

RESULT 4  
195746 878 bp DNA PAT 17-JUL-1998  
LOCUS Sequence 8 from patent US 5733748.  
DEFINITION  
ACCESSION 93940216  
NID 195746.1 GI:3940216  
VERSION  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 878)  
Yu,G. and Rosen,C.  
TITLE Colon specific genes and proteins  
JOURNAL Patent: US 5733748-A 8 31-MAR-1998;  
FEATURES  
source  
1..878  
/organism="unknown"  
BASE COUNT 257 a 179 c 188 g 241 t 13 others  
ORIGIN

Query Match 72.5%; Score 76.8; DB 5; Length 878;  
Best Local Similarity 91.0%; Pred. No. 2.7e-12;  
Matches 81; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GGCATTCACATTTTAAATATATGGAAGTGGATAGAGACGCGCTGCAATAGCC 60  
DB 717 GGCATTCACATTTTAAATATATGGAAGTGGATAGAGACGCGCTGCAATAGCC 776

QY 61 TAGGCGTGAATTTTGTCTGCAATATAA 89  
 ||||| ||||| ||||| ||||| |||||  
 Db 777 TAGGGGTGAATTTTGTGGGTGAATATA 805

RESULT 5  
 AB017156 2937 bp mRNA ROD 04-MAR-1999  
 LOCUS Mus musculus gob-5 mRNA, complete cds.  
 DEFINITION AB017156  
 ACCESSION 93721911  
 MID AB017156.1 GI:3721911  
 VERSION GOB-5.  
 KEYWORDS Mus musculus adult intestine goblet cell cDNA to mRNA.  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (sites)  
 AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
 TITLE Cloning and identification of the gene gob-5, which is expressed in  
 intestinal goblet cells in mice  
 JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
 MEDLINE 99160866  
 REFERENCE 2 (bases 1 to 2937)  
 AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru  
 Komiyama, ERATO, JST, Hirohashi Cell Configuration Project, 5-9-9,  
 Tokodai, Tsukuba, Ibaraki 300-2633, Japan  
 (E-mail:tkomihcpc.jst.go.jp, Tel:81-298-47-7563,  
 Fax:81-298-47-5226)

FEATURES  
 source  
 Location/Qualifiers  
 1..2937  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_type="goblet cell"  
 /dev\_stage="adult"  
 /tissue\_type="intestine"  
 15..2756  
 /gene="gob-5"  
 15..2756  
 /codon\_start=1  
 /protein\_id="BA33743.1"  
 /db\_xref="PID:01034712"  
 /db\_xref="PID:93721912"  
 /db\_xref="GI:3721912"  
 /translation="MESLSPVFLILHLGVLSESLIOLNNNGYEGIVAIIDHVP  
 EDVALIHIKIMVQAAPYLFEANGKRRYFKNVALIPESKAKDEYRRLTEPKNA  
 DVLVSTSPINDPEYTHIHGACGKGRIRHLTPPLAGKLTQYGPDRFTVEKRAH  
 FRWGFENYNDEKFTLSKGRQAVRCSAALTGNQVRCCGGSCITNGKVIDRVTG  
 LYKNCVPEPDHONKASIMFNONINSVECTEKNQAPNDONQNCMLRSTWEY  
 IOESDFEKPMTAPQAPFSLQIQRIQVCLVLDKSGMLNDNRNNAQSLRT  
 LLOVQSGWGMVTFDSAAVQSELKQNSGADRLILKHLPTVSAGTSCISLRT  
 AFTYKKKPPDGSSEIVLLDGEENLTSSCRDIYKQSAIIHVAIGPAAKGELOS  
 KATGGLOITSSDQVONGLVAFALSSGNALIAQHSIQLESRCYNLONNMWMSVI  
 VDSSVGDITLFLITWTHRPPIFLIWDPSGVQNGILDTTKVAILOVPGTAKGVFW  
 YSIQASQTLTLITSRASATLPIITVPVNNKGTGFPSPVTVYASIRGASPIELK  
 ASVTALIESVNGKIVTLELDNGADATKNDGVSRFTAFDANGRSVKIMLGVY  
 TSDORAPAPKNRAVYIDGWIEDGEVRNPNRPETSYVDOLCFSTSGGSFVATN  
 VPAAPRIPDLFPPCOITDLKASIOGQNTVNTMTAPGDDYDHGRASNTIIMSTSYD  
 LRDHNTSLQVNTGILPKKASSEEIIEFEGNGTENGTDIFAIQIVDSDNLSSEI  
 SNIAKVSFIIAOPPEIPEDSTPPCDIINSTIIPGIVHLKIMMKWLGEMOVTLGLH"

BASE COUNT 860 a 718 c 693 g 666 t  
 ORIGIN

Query Match 44.7% Score 47.4; DB 12; Length 2937;  
 Best Local Similarity 70.8%; Pred. No. 0.00041;  
 Matches 63; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 GGCAATCCATTTTAAATATATGAGAGTGAATGAGAACTGCAGCTGCATAGCC 60  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 2688 GGCAATCCAGCTGCTGAAGTAATATGAGAGTGGCTAGGGGAATGCAGCTAGCT 2747  
 QY 61 TAGGCGTGAATTTTGTCTGCAATATAA 89  
 ||||| ||||| ||||| ||||| |||||  
 Db 2748 TTGCACCTGAATTTTACGACAGAAATCAA 2776

RESULT 6  
 BS16829KB 29324 bp DNA BCT 20-NOV-1997  
 LOCUS Bacillus subtilis 29kb DNA fragment from ykwc gene to cse15 gene.  
 DEFINITION AJ222587  
 ACCESSION AJ222587  
 MID 92632216  
 VERSION AJ222587.1 GI:2632216  
 KEYWORDS Chev gene; cse15 gene; kina gene; pata gene; ykua gene; ykub gene;  
 ykuc gene; yukd gene; ykue gene; yukf gene; ykug gene; yukh gene;  
 yukl gene; yukp gene; yukr gene; ykum gene; yukn gene;  
 yuku gene; yukv gene; yukw gene; ykwc gene; ykwd gene;  
 Bacillus subtilis.  
 SOURCE Bacillus subtilis  
 ORGANISM Bacillus subtilis  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
 Bacillus.  
 REFERENCE 1 (bases 1 to 29324)  
 AUTHORS Scanlan, E. and Devine, K.M.  
 TITLE Sequence of the Bacillus subtilis chromosome from ykua to cse-15  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 29324)  
 AUTHORS Devine, K.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-1997) K.M. Devine, Genetics, University of  
 Dublin, Trinity College, Lincoln Place Gate, Dublin 2, Ireland

FEATURES  
 source  
 Location/Qualifiers  
 1..29324  
 /organism="Bacillus subtilis"  
 /strain="168"  
 /db\_xref="taxon:1423"  
 89..955  
 /gene="ykwc"  
 89..955  
 /gene="ykwc"  
 /note="homologous to 3-hydroxyisobutyrate dehydrogenase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ykwc protein"  
 /protein\_id="CA10859.1"  
 /db\_xref="PID:e1181901"  
 /db\_xref="PID:92632217"  
 /db\_xref="GI:2632217"  
 /db\_xref="SWISS-PROT:O34948"  
 /translation="MKRTTIGFTIGLVGKSMASHIINDGHPVLVYTRKKAESILQK  
 GAIKKDVKLSKEDAVIITMVGPSDVEYFESNGIENAKGAYLIDMTSKPRL  
 AKAIAEAKKALEFALDAPVSGGICGQNTLAIWVGEEAEAFACAPRISLMEENTQ  
 YOGAPAGQGHKQKNOVAIAAGMVGVAEAAVQKSGLEPENYVKTSTTGAGGWSIS  
 NIAEPMQOGHPEPFFYKHKIKMGIGALIEAEALGEMPELSAKSLIDYLAANGSEN  
 SGDSITKLVK"

complement(94..1767)  
 /gene="ykwd"  
 complement(94..1767)  
 /gene="ykwd"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ykwd protein"  
 /protein\_id="CA10860.1"  
 /db\_xref="PID:e1181902"  
 /db\_xref="PID:92632218"  
 /db\_xref="GI:2632218"  
 /db\_xref="SPRMBL:O31398"  
 /translation="MKRAFLISAAGAGLFTFGVQOASAKESCPQVTVTKGTNQ  
 NMSLNDVAKLHINTNITLANAEKEMKLOLKHAKOSVAVKQVQKTEATAPQKT  
 TEKAADQNTASKAPATAEKNTTTSPPSSVSAVEKVEELTNAEROKGGLKPLQIDE  
 TISAKASQDMKDKVDPHOSPTVSSPDMKSPSISYKTAGENIAKOKTPEEVY  
 KAMNSGCHRNILNPNFTIHIGVYVSSGIWTOQFIK"

```

gene
2104. 4218
/gene="ykua"
CDS
/note="homologous to penicillin binding protein 2B"
/codon_start=1
/transl_table=11
/product="ykua protein"
/protein_id="CA10861.1"
/db_xref="PID:e1181903"
/db_xref="PID:g2632219"
/db_xref="GI:2632219"
/translation="MTETGRPEPKKSKGRRAIRNMLFLIAYEYVLTALFKLGVOIV
EEQHEDEKSNKATAYYPARGKAYDNQAVADNOSVPELIVYSTSTEDIK
TAKLASFHIDTEFLERDLRDYWAAPKKAALLKESNLKGDQAYKQIERVP
DOELRAIQODDEEMETAAIYTRFSSGNAYEPOIVKAMPNKSNGKNGALDEKNS
SORPNKDLTYDEISIVSEHLEELPGIDYNDMTRKYPDKTLYSVFGVTPDGLS
DKRDYITRGVANNDRVCKSYLEYOEELNHSKEVEVEDNKGAVSOKTIDKGR
GYDLOSFDMELQAKYEKILIEEVRNSRARGNIMDRAPVAMDPNNGDILSNAGKI
DLTKNKIEDYAIQAFITQYEMGSANVAGAVLAGYDGIPIRYIIDAPMLGTNLTK
SYTNMGTINELTALOKSSNVNFMVAMHLAGVYKPHGSLPADNDLKMRYVSQFG
LGVKTGIDLPQESAGMOTPKYVGLIDLAIGQDXTYLPADQAYISYIANGYVVO
PRIVSIEHKPKKDLGKAIIEHRKPKVLKKNINSEDLKOVOTGMKLYTSSGAKTFF
TEDVSGKTCTATFTFYGTNRNMGKTTYMLTGVGYPSKRPVAVSVVPSYDDDKI
NTIARKAIHATAELEKHSKR"
4382. 6202
/gene="kua"
4382. 6202
/gene="kua"
CDS
/function="sporulation kinase"
/codon_start=1
/transl_table=11
/product="kua protein"
/protein_id="CA10862.1"
/db_xref="PID:e1181904"
/db_xref="PID:g2632220"
/db_xref="GI:2632220"
/translation="MEODTOHYKPLQTKTDIHAVLASNGRIIYISANSKILHLYGGE
MGSFLKTFLEHEDOFVSEYVNEHMLKCFERFKDHTIYVWEAAVEAYITRBR
TERELIYKMKVLEEFEGHOSLNCERHEIERPASESTTYITDOYERVENLEPSCISY
KKRIYVNSNMLSVGAKSKDAIIGSSSTFEIEERHDIVYKNRIITMOGMEYAMIEO
TWKRLDGTVPVLEKVASPTVYKNOQALLLIDISSRKFOTILKSRERYQLLQNS
IDITIAVHNGKSFVFNESGISLEAATYEDLDGKNYVDLHPCDHEDVERLONTAEQ
KRESEIVOSKFTFORVITYTEWVCIPTFEGEAAVOVLIDISEKROEEMLKAEK
LSIAGIAGIAHEIRNPLTAKIGFOLKPMEGNEHYFDIYFSELSLIELLSFL
MAKPOQNAVREYLNKILIGESVALLFQANLNGCFIRTSKSDSYINGQONOLK
VTINIKRNVESAPDGGIVDITITDEHSVHTVYKREGIPEKVLNRIGEPFLITKE
KOTGLGIAVTFNIENHOGVINDSHPEKGAISFPK"
complement(6213. 7391)
/gene="pata"
complement(6213. 7391)
/gene="pata"
/note="hisc homologue"
/codon_start=1
/transl_table=11
/product="Pata protein"
/protein_id="CA10863.1"
/db_xref="PID:e1181905"
/db_xref="PID:g2632221"
/db_xref="GI:2632221"
/translation="MEHLNPKAREIEIGIRKFSNLAHQHEDVLSITIGOPDEFTH
HYKAAAKKAIIDENVTSTPNAGLELRQAVOLYMKKADPNDASEIITITGAOIH
AAFTLISPGDEVIMGPYVPGYEPILNCGKRPVYVSDTSEFKTLAAILADLPN
TKCVYLPYPSNPTGVTLSSEELKSTIALKNGNVYLSPEIYSELTYDRPHYSIATY
RQOTIVINGLSKSHSTGWRIGLELPAPKIDIAHILKLVHYNVSCASSISQKALEVY
NIDIDALINREQYKRLDYVRLVMSGLDVYKPGSAGTIPFSIKFSGMTSPDSVAL
LEDAGVALVPGSSFSYEGEYVRLSFACSMDTLREGIDRLLEFLVKKREAMOTINNGV"
7958. 8869
/gene="chev"

```

```

CDS
7958. 8869
/gene="chev"
/function="Involved in chemotaxis"
/codon_start=1
/transl_table=11
/product="Chev protein"
/protein_id="CA10864.1"
/db_xref="PID:e1181906"
/db_xref="PID:g2632222"
/db_xref="GI:2632222"
/translation="MSIOVEYELDSGNELETVKFGVGENAFGINVMVREILPOVE
VYSQIHRVSWENAIETKPTSLNQGEMERHLGILKLEDMATLPYERKAITIDIESGVDI
YNMHTEGDERTDKRLIIVEDSPILMLRLLOELKAGVNNIASENGEAEYITMLT
AENEDLSKQIMIDITDIEMPMDCHRLTKLKENPKSSDVPVMISSLITDLNHRG
EYVGADEQISKEISDLIKKVDYVIE"
complement(8913. 9377)
/gene="ykub"
complement(8913. 9377)
/codon_start=1
/transl_table=11
/product="ykub protein"
/protein_id="CA10865.1"
/db_xref="PID:e1181907"
/db_xref="PID:g2632223"
/db_xref="GI:2632223"
/db_xref="SPTREMBL:034548"
/translation="MDHAYTKDLOPTYENLSKAVYTVNRHAKTAPNPKYLLKKRA
LOKLYREGKAKIGLHFSKNRPFSSQOSDYLISIDYTRHNPTEDEPHDLHGLTN
OSYRNPKAOMSLFKAKHLQETVGMKEKPLVNRQOPAYHAKFVFKLGESTY"
complement(9503. 10795)
/gene="ykuc"
complement(9503. 10795)
/note="homologous to macrofide efflux determinants"
/codon_start=1
/transl_table=11
/product="ykuc protein"
/protein_id="CA10866.1"
/db_xref="PID:e1181908"
/db_xref="PID:g2632224"
/db_xref="GI:2632224"
/translation="MDIFKNRNFVLEFPAALASOMGTTVGNNAFAFFLIDRESSOPSY
TTLAELMSLPTIIEFLIYGVADDEFDRKRVENCDWIRAGLTVLFLFLFNGNIPLV
RQILFIRSAVYKFFPPAPNSIYOALIPREHYKAGLNGMLESITMVEGVGAGMYN
TIGIGAILLIDVSEITISGLIRSCRIRKAPNGARNSWRTSYKDSINDREGILY
IKNNLLASLIEGFTIFGVNGGFVVLPEFTMKYGLADPRYEMHVSFTIALGFIILA
GSVITLISKRVKPHFLMSIPFIAGLLEFYIGYNNILMVYVAAAFALGMCIGPINIA
IGGMPKIVHPKLMGRVSGMDPFMMFQOSLTLGLVALLFPFVSNIDYLYXGMGVII
LLVFIENFALPKYSAGQAVENVOEAFQDQPKKARSV"
complement(10871. 11365)
/gene="ykud"
complement(10871. 11365)
/gene="ykud"
/codon_start=1
/transl_table=11
/product="ykud protein"
/protein_id="CA10867.1"
/db_xref="PID:e1181909"
/db_xref="PID:g2632225"

```

```

Query Match      34.5%  Score 36.6; DB 1; Length 29324;
Best local similarity 60.6%  Pred No. 0.4;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 1 GGCATTACATTTAAATTAATGATGATAGAGAACTGACGCTGCAATAGCC 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7888 GCATTCACATTTTACAAATTAACCGCATATATACGATCAATGCAATAGCG 7947
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 61 TAGGCGTGAATTTTGTCCAGATTAATAAATAATCATTT 99

```



SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

Bacillus subtilis.  
Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
Bacillus.  
1 (bases 1 to 208230)

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,  
Azevedo, V., Bertolo, M.G., Bessières, P., Bolotin, A., Borchert, S.,  
Boriss, R., Boursier, L., Brans, A., Braun, M., Briganti, S.C.,  
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,  
Carter, N.M., Choi, S.K., Codani, J.J., Conneron, I.F., Cummings, N.J.,  
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,  
Emmerson, P.T., Entlian, K.D., Errington, J., Fadrer, C., Ferrari, E.,  
Foulegre, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Gallizzi, A.,  
Galleron, N., Ghm, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,  
Grandi, G., Giuseppe, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,  
Henaui, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,  
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,  
Klaerner-Bianchard, M., Klein, C., Kobayashi, Y., Koetter, P.,  
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,  
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,  
Masuda, S., Meisel, C., Medigue, C., Medina, N., Mellado, R.P.,  
Mitsuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,  
Ogawa, K., Ogihara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,  
Portelle, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P.,  
Punelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,  
Rivolta, C., Roche, E., Roche, B., Rose, M., Sadale, Y., Sato, T.,  
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,  
Sekowska, A., Seror, S.J., Serro, P., Shin, B.S., Soldo, B.,  
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,  
Takeuchi, M., Tamakoshi, A., Tanaka, T., Teptstra, P., Tognoni, A.,  
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,  
Viari, A., Wandup, R., Wedler, E., Wedler, H., Wellenreger, A.,  
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,  
Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and  
Danchin, A.

The complete genome sequence of the gram-positive bacterium  
Bacillus subtilis  
Nature 390 (6657), 249-256 (1997)

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 208230)  
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.  
Direct Submission  
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,  
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724  
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,  
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45  
68 89 48

FEATURES  
SOURCE

Location/Qualifiers  
1..208230  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
46..183  
/gene="ykdZ"  
46..183  
/gene="ykdZ"  
/function="unknown"  
/note="similar to hypothetical proteins from B. subtilis"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB13202.1"  
/db\_xref="PID:e1184919"  
/db\_xref="PID:e1184919"  
/db\_xref="GI:2633700"  
/db\_xref="GI:2633700"  
/db\_xref="SPTREMBL:O34405"  
/translation="MEKKKEQYINQAEVPHPTKGEYALPLHETHTLSEDEDTQT  
E"  
689..2044  
/gene="ykoK"  
689..2044  
/gene="ykoK"  
/function="unknown"  
/note="similar to Mg2+ transporter"

gene  
CDS

/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB13202.1"  
/db\_xref="PID:e1184920"  
/db\_xref="PID:e1184920"  
/db\_xref="PID:e1184920"  
/db\_xref="GI:2633701"  
/db\_xref="GI:2633701"  
/db\_xref="SPTREMBL:O34442"  
/translation="WQNMVYDELILRIIILLRDKGIRDFRSVIDELQPYMAFIKFE  
MPEHRARVLYLVDDITDMIGELERFOLVAVNKVGTAKATLANKMDNDLAQL  
EEMBELEKQLSMEASESKAVOLINMPADSAGRMATNRYVWIPQHYKDAVVK  
KSPLEIASINYLIVNESKOLGVLSYRDLIGEPPEKVODLMEFRVYSADADODE  
EVALIRYDRIAPVVENNVIGYIVDDIIVVIREADEVDKPKASGKDIPTD  
KATVAAVRRPLMLLLFTGISSSIISYFEDALKOVYALFPPPMNGMNGTOS  
LAIVRIGSKSEMKKTIVRIIFRPFRTSIFIGAVCSVLIVASTIIVGNNLGVVA  
SLPLTLLIGTMSGIIPIIILHKLKVPALASGPLITLNDLISLIYFGIATAFHS  
I"  
complement(2087..2419)  
/gene="tnrA"  
complement(2087..2419)  
/gene="tnrA"  
/function="global nitrogen regulation (positive regulation  
of nrgAB, nasB, gabP, ureABC; negative regulation of  
glnR)"  
/note="alternate gene name: scgR"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="CAB13204.1"  
/db\_xref="PID:e1184921"  
/db\_xref="PID:e1184921"  
/db\_xref="GI:2633702"  
/db\_xref="SPTREMBL:O45666"  
/db\_xref="SPTREMBL:O45666"  
/translation="MTEDSHYKQKVIISIGIVSELTGLSVQRIYVEERLIYQPS  
SRGRTKSYFAVDERLMDIANKREDGVQAEILKMRKKEQMLKNDPQVRKKMLEQOLN  
AHPFRKNR"  
2614..2769  
/gene="ykbZ"  
2614..2769  
/gene="ykbZ"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB13205.1"  
/db\_xref="PID:e1184922"  
/db\_xref="PID:e1184922"  
/db\_xref="PID:e1184922"  
/db\_xref="GI:2633703"  
/db\_xref="SPTREMBL:O34923"  
/translation="MEKKKAVSLAKKEIILEDIKRDEMLETMQALGEAFQPLNSYVN  
GQYRKS  
2857..3039  
/gene="ykoL"  
2857..3039  
/gene="ykoL"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB13206.1"  
/db\_xref="PID:e1184923"  
/db\_xref="PID:e1184923"  
/db\_xref="PID:e1184923"  
/db\_xref="GI:2633704"  
/db\_xref="SPTREMBL:O34763"  
/translation="WSNLKSLAKERKRHHSEKLYQIGVYKNEVNNKMTISELKREYA  
YFPRSTNHNKYPYTR"  
3172..3666  
/gene="ykoM"  
3172..3666  
/gene="ykoM"  
/function="unknown"  
/note="similar to transcriptional regulator (MarR family)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB13207.1"  
/db\_xref="PID:e1184924"  
/db\_xref="PID:e1184924"

```
/db_xref="PID:g2633705"
/db_xref="GI:2633705"
/db_xref="SPREMBL:O34949"
/translation="MMRLSENEEVERAMNLYRVARAFKSVSEHSINDSKHGNPT
EFAVELLYTRGPOLQOIGSRLLVSGNTVYVDIKLERNGFLVREODPKRKYVAH
LTDGNEYLDKIYRIHMLRIRARFSGSPDQDLIVLKKAGIHSOHLER"
3645..3666
/gene="ykop"
/complement(3651..4772)
/gene="ykop"
/complement(3651..4772)
/gene="ykop"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/translation="MKNLIPPLSISTGHHNADALQAELESOGLAERKIDIFSHY
RLEKLSVAYLKWIOYEPKTSIGIYRLLAGEFQHDKRYMECEVFOQRHLOEK
OPDIACFTHVAIYLLNLRKPEYPMIVYVYTDPEVQMLGRKNDIHYEPSTEVK
OLISEGIDONNIVYLGIDPHONFEMESADTQHHPPYITITGSMGVGILKVVOL
SPGKILYKICGRNEKLYSVKSLHPLIARLYLHKAEMNLREBQATIMKRPQS
VTISECLOKRLPVFIYHALPGQSENNLLERKLVTDNRWMDQKAEYIAAFQSN
EOMREYKRVHVGVIYGEUSDRIKIVLKRIWKORNTLTK"
4864..5415
/gene="ykop"
/translation="MRSVAVFFPALPESILYSSNNKPEFWILLERKROFLYKSEF
SKSNLTAVINYLDELARKGVPEASLYLVAAGMSNALILEKLNKESDERVYLLD
ESPEKTLFSLAKWALENNYVDITDINLTKFSKKNVSYHDOSGSEMSNNE
LMSFYTITSLKQYRGAFGIDFTFNRLFRFNRYVIDTHQLSLVYLHAKYEISTHG
LIAVAVGILTRMDYVOCYEPYQFNKRSSVQKINKKAKRNKAFESDFEIVNVA
ATYVNSNGNYKFRADITWKLVDLAPRESEYIAKRSQINSVLYLKN"
complement(1156..1590)
/gene="BB0090"
/complement(1156..1590)
/gene="BB0090"
/note="similar to GB:L77117 SP:Q57674 PID:1590959 percent
identity: 32.61; identified by sequence similarity;
putative"
/codon_start=1
/translation="v-type ATPase, subunit K, putative"
/product="v-type ATPase, subunit K, putative"
/protein_id="AAC66487.1"
/db_xref="PID:g2687983"
/db_xref="GI:2687983"
/translation="MDIGLIGVNSALTISGSLGAGGAAGSAIGAKRKYOGKPA
PFLILVPSAPINQIYGYITLMTNLYVNMOTPMILIGIGGGAFAINAGRAOGKA
AAGACDAFSETGKGFAIYLLVGLIESVALFVWFMTIKFY"
complement(1607..3433)
/gene="BB0091"
/complement(1607..3433)
/gene="BB0091"
/note="similar to GB:D17462 GB:x76913 SP:PA3439 PID:472918
PID:487273 percent identity: 27.18; identified by sequence
similarity; putative"
/translation="v-type ATPase, subunit I, putative"
/product="v-type ATPase, subunit I, putative"
```

```
REFERENCE
AUTHORS
1 (bases 1 to 12984)
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
Gocayne,J.D., Weidman,J., Uterback,T., Matthey,L., McDonald,L.,
Artach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi
JOURNAL
NATURE 390 (6660), 580-586 (1997)
MEDLINE
98065943
2 (bases 1 to 12984)
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
Gocayne,J.D., Weidman,J., Uterback,T., Matthey,L., McDonald,L.,
Artach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
Direct Submission
Submitted (12-DEC-1997) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..12984
/organism="Borrelia burgdorferi"
/db_xref="taxon:139"
complement(110..1084)
/gene="BB0089"
complement(110..1084)
/gene="BB0089"
/note="hypothetical protein: identified by Glimmer;
putative"
/codon_start=1
/product="B. burgdorferi predicted coding region BB0089"
/protein_id="AAC66480.1"
/db_xref="PID:g2687986"
/db_xref="GI:2687986"
/translation="MRSVAVFFPALPESILYSSNNKPEFWILLERKROFLYKSEF
SKSNLTAVINYLDELARKGVPEASLYLVAAGMSNALILEKLNKESDERVYLLD
ESPEKTLFSLAKWALENNYVDITDINLTKFSKKNVSYHDOSGSEMSNNE
LMSFYTITSLKQYRGAFGIDFTFNRLFRFNRYVIDTHQLSLVYLHAKYEISTHG
LIAVAVGILTRMDYVOCYEPYQFNKRSSVQKINKKAKRNKAFESDFEIVNVA
ATYVNSNGNYKFRADITWKLVDLAPRESEYIAKRSQINSVLYLKN"
complement(1156..1590)
/gene="BB0090"
/complement(1156..1590)
/gene="BB0090"
/note="similar to GB:L77117 SP:Q57674 PID:1590959 percent
identity: 32.61; identified by sequence similarity;
putative"
/codon_start=1
/translation="v-type ATPase, subunit K, putative"
/product="v-type ATPase, subunit K, putative"
/protein_id="AAC66487.1"
/db_xref="PID:g2687983"
/db_xref="GI:2687983"
/translation="MDIGLIGVNSALTISGSLGAGGAAGSAIGAKRKYOGKPA
PFLILVPSAPINQIYGYITLMTNLYVNMOTPMILIGIGGGAFAINAGRAOGKA
AAGACDAFSETGKGFAIYLLVGLIESVALFVWFMTIKFY"
complement(1607..3433)
/gene="BB0091"
/complement(1607..3433)
/gene="BB0091"
/note="similar to GB:D17462 GB:x76913 SP:PA3439 PID:472918
PID:487273 percent identity: 27.18; identified by sequence
similarity; putative"
/translation="v-type ATPase, subunit I, putative"
/product="v-type ATPase, subunit I, putative"
```



QY 5 TTCACATTTTAAATTTATGTCGAGTGCAGTAAGTGCAGCTGCATACGCTAG 64  
Db 54556 TTTTATTTGAAAAATTTGTCGTCACCTGACGAGAAAAAGCTGATGACAAAGACGCGA 54615  
QY 65 GCTGAATTTTGTGTCAGTAATAATAATAATCATTCAT 102  
Db 54616 AGAATATTGTCAGCAGACGAAACATTAAGTAATGTTCTT 54653

RESULT 12  
CEY7A9 224129 bp DNA HTG 03-DEC-1998  
LOCUS CEY7A9  
DEFINITION Caenorhabditis elegans chromosome IV clone Y7A9, WORKING DRAFT  
SEQUENCE, in unordered pieces.  
ACCESSION 295312  
NID 93355732  
VERSION 295312.1 GI:3355732  
KEYWORDS HTG: HTGS, PHASE1.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 224129)  
AUTHORS McMurray, A.  
JOURNAL Direct Submission  
Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridgeshire CB10 1HQ, UK and Department of Genetics,  
Washington University, St. Louis, MO 63110, USA. E-mail:  
jessanger.ac.uk or rwenematode.wustl.edu  
On Jul 30, 1998 this sequence version replaced gi:2528903.  
COMMENT Order of segments is not known: 800 n's separate segments.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E. coli, yeast, vector,  
phage etc.  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES  
source 1..224129  
Location/Qualifiers  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="IV"  
/clone="Y7A9"

BASE COUNT 72928 a 37724 c 37595 g 73444 t 2438 others

ORIGIN

Query Match 32.1%; Score 34; DB 34; Length 224129;  
Best Local Similarity 59.2%; Pred. No. 2;  
Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 5 TTCACATTTTAAATTTATGTCGAGTGCAGTAAGTGCAGCTGCATACGCTAG 64  
Db 121820 TTTTATTTGAAAAATTTGTCGTCACCTGACGAGAAAAAGCTGATGACAAAGACGCGA 121879  
QY 65 GCTGAATTTTGTGTCAGTAATAATAATAATCATTCAT 102  
Db 121880 AGAATATTGTCAGCAGACGAAACATTAAGTAATGTTCTT 121917

RESULT 13  
CEY7A9C 26821 bp DNA INV 23-NOV-1998  
LOCUS CEY7A9C  
DEFINITION Caenorhabditis elegans cosmid Y7A9C, complete sequence.  
ACCESSION 299286  
NID 92414426  
VERSION 299286.1 GI:2414426  
KEYWORDS HTG.

SOURCE  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 26821)  
AUTHORS McMurray, A.  
JOURNAL Direct Submission  
Submitted (17-SEP-1997) Louis, MO 63110, USA. E-mail:  
jessanger.ac.uk or rwenematode.wustl.edu  
2 (bases 1 to 26821)  
REFERENCE Wilson, R., Alnsough, R., Anderson, K., Baynes, C., Berks, M.,  
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,  
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,  
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,  
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,  
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,  
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,  
Saunders, D., Showkeen, R., Smalton, N., Smith, L., Sonhammer, E.,  
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,  
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,  
Wilkinson-Sproat, J. and Wohlman, P.  
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans  
Nature 368 (6466), 32-38 (1994)  
http://webc.sanger.ac.uk/cgi-bin/displaydb-wormscsclass-sequence/subject-Y7A9C  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.  
IMPORTANT: This sequence is not the entire insert of clone Y7A9C.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone C11P10 is at 1 in this sequence. The  
true left end of clone K03D3 is at 26718 in this sequence. The true  
right end of clone F38C2 is at 7539 in this sequence. The start of  
this sequence (1..104) overlaps with the end of sequence 282267.  
The end of this sequence (26718..26821) overlaps with the start of  
sequence 282276.

FEATURES  
source 1..26821  
Location/Qualifiers  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="IV"  
/clone="Y7A9C"  
complement(5759..9680)  
/gene="Y7A9C.3"  
complement(join(5759..6305,6360..6469,7606..7798,  
7846..7961,8956..9108,9168..9314,9434..9472,9516..9680))  
/gene="Y7A9C.3"  
/codon\_start=1  
/protein\_id="CAB16543.1"  
/db\_xref="PID:e1350804"  
/db\_xref="PID:g3881322"  
/db\_xref="GI:3881322"  
/db\_xref="SPTREMBL:O18258"  
/translation="MDLSKLNNEFEKMGOTATGILIVPGITILLVLAIPPVY  
VRLNRRDNRKIVLPVNVHFFQMDTPPSHYDSDSFYNYIIPLKEIYLVITYSN  
INSSRRINFGLSMIIVAFVALLFFSTFETIPIMLSAKMSAONCAOVAKITVOT  
LIVFVCKSLPFTVSHFEVMEVAFYLVLLIICIPILFVATHFSIAVYIVICLAF  
FIILAIYOFETREHFIIFLLAORFLVFPSPTEKHIALPQKFNKSVWVYHLLTLP



TGVHYNNVATITFEYLFVSFIICMLSPNNGPELSNLILFIYALITPALTSTAEAFH  
 ILFIILAAORFLVFEFSTPEKHIALFOKEMSKRWVYVHLVSVEKTEITICIVSSQMEIK  
 RPNINCIIPAGICODTWSGICGTYTVMWMTLLFSTKIKLOSLIFPHLFLSTIPFY  
 IPIMISAOKISSNOYCKVQNYIMWQTLITFCKLVQHOF"  
 12999..17050  
 /gene="x79ac.1"  
 join(12999..13077,13538..13741,14804..15408,16057..16135,  
 16188..16393,16589..17050)  
 /gene="x79ac.1"  
 /note="predicted using GeneFinder; similar to zinc finger,  
 C3HC4 type (RING finger)"  
 /codon\_start=1  
 /protein\_id="CAB16542.1"  
 /db\_xref="PID:e1350803"  
 /db\_xref="PID:g3881321"  
 /db\_xref="GI:3881321"  
 /db\_xref="SPTREMBL:O18257"  
 /translation="MFHLSLFCGCKPFLGFEERHHPYDNDKRRICSSGSGSLCK  
 KCFERYKNCPIAKPLVRDNDMAREIYAFLEKTPHIVFDMMNNSQYQKNTDLOK  
 ELIKVSKERFRFGILANRYICSCDFLEHAGRAVIFLDEMISALFEISDSATHVO  
 FLMEIEIERGRKCLRTMRKRAFLKLEYLACDSDLDDYDLPKFEFLDMKVPD  
 GMEKVRKNLEHOFASANYPWDCRCRYAMRIMDMFSKNOIEFLVSIVNDLHE  
 ITECFKNSQITDMLQTKRMVDIOLKLYTLFCWNNYKDGVELEFNNSNRLPKCCL  
 CKMSTCAECWDEEPKCGICDDATSSRPDIIVLIELYKHKENCISLEAMECCDA  
 AOLGCLNCASITDELETCYVELYVERSALETPEVTSNEFTKMDQLRTTYGSGFP  
 NMWQSDCTTRHAKPTADWKNCGNHFRRSMYASVSSCFEHNHKSPLRLKEIKNY  
 DLMKRAAFGLVFKIKIEKIKHVESEHAKKEILET"  
 BASE COUNT 8877 a 4503 c 4631 g 8810 t  
 ORIGIN  
 Query Match 32.1% Score 34; DB 36; Length 26821;  
 Best Local Similarity 59.2% Pred. No. 2.1;  
 Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
 QY 5 TTCACATTTTAAATTTATGTGGAAGTGGATAGCAAGTGCAGCTGTAATAGCTAGG 64  
 DB 11587 TTTTATTTGAAAATTTGTGTACTTACCTCGACAAAAGCTGATGACAAAGACCGA 11646  
 QY 65 GGTGAATTTTTCAGATAAATTAATATCATTCAT 102  
 DB 11647 AGAATATTGTGACAGACATTAAGTAAATGTTCTT 11684  
 RESULT 14  
 CER28F4 30567 bp DNA INV 23-NOV-1998  
 LOCUS Caenorhabditis elegans cosmid T28F4, complete sequence.  
 DEFINITION 272817  
 ACCESSION 91301733  
 NID 272817.1 GI:1301733  
 VERSION HTG: Mechanosensory protein 1like; Protein-tyrosine phosphatase.  
 KEYWORDS Caenorhabditis elegans.  
 SOURCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;  
 Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peioderinae; Caenorhabditis.  
 ORGANISM  
 REFERENCE 1 (bases 1 to 30567)  
 AUTHORS McMurray, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAY-1996) Louis, MO 63110, USA. E-mail:  
 jesssanger.ac.uk or jwemmatode.vustli.edu  
 2 (bases 1 to 30567)  
 WILSON, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,  
 Bonfield, J., Burton, M., Connell, M., Copsey, T., Cooper, J.,  
 Coulson, A., Craxton, M., Dear, S., Du, Z., Durlin, R., Favell, A.,  
 Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,  
 Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,  
 Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,  
 O'Callaghan, M., Parsons, J., Percy, C., Rikken, L., Roopra, A.,  
 Saunders, D., Showkhen, R., Smaildon, N., Smith, A., Sonhammer, E.,  
 Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,  
 Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,  
 Wilkinson-Sproat, J. and Wohlman, P.

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans  
 Nature 368 (6466), 32-38 (1994)  
 94150718  
 Coding sequences below are predicted from computer analysis, using  
 predictions from GeneFinder (P. Green, U. Washington), and other  
 available information.  
 For a graphical representation of this sequence and its analysis  
 see: "  
 http://webace.sanger.ac.uk/cgi-  
 bin/display?db=wormaceclass=sequence&object=T28F4  
 Current sequence finishing criteria for the C. elegans genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 IMPORTANT: This sequence is NOT necessarily the entire insert of  
 the specified clone. It may be shorter because we only sequence  
 overlapping sections once, or longer because we arrange for a small  
 overlap between neighbouring subclones.  
 IMPORTANT: This sequence is not the entire insert of clone T28F4.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we arrange for a small overlap between  
 neighbouring subclones.  
 The true left end of clone T28F4 is at 1 in this sequence. The true  
 right end of clone T28F4 is at 10921 in  
 sequence 722503.  
 The true left end of clone C26C6 is at 30464 in this sequence. The  
 true right end of clone ZK524 is at 7785 in this sequence. The  
 start of this sequence (1..104) overlaps with the end of sequence  
 723912.  
 The end of this sequence (30464..30567) overlaps with the start of  
 sequence 722503.  
 Location/Qualifiers  
 1..30567  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="1"  
 /clone="T28F4"  
 /gene="T28F4"  
 complement(7250..8389)  
 /gene="T28F4.1"  
 complement(join(7250..7408,7639..8203,8256..8389))  
 /note="T28F4.1"  
 /note="predicted using GeneFinder; cDNA EST yk196c3.3  
 comes from this gene; cDNA EST yk196c3.5 comes from this  
 gene; cDNA EST yk426d11.3 comes from this gene; cDNA EST  
 yk426d11.5 comes from this gene; cDNA EST EMBL:T00687  
 comes from this gene; cDNA EST EMBL:T00688 comes from this  
 gene"  
 /codon\_start=1  
 /protein\_id="CA96693.1"  
 /db\_xref="PID:e1349707"  
 /db\_xref="PID:g3880343"  
 /db\_xref="GI:3880343"  
 /db\_xref="SPTREMBL:Q22849"  
 /translation="MNMTYKLNLDLIEIFGLCNVALHKKCSKSEKKRAGCGGAMRK  
 RYLIRKFNVSDFKMPKRTSSSRSDMASDSEYEYDMSDVSFDEDEDERLIDLTH  
 YEHAHKNQNDLMLHQPYGTDMGFGGSSMSMASAGMTSGYSASASVFG  
 YQPDWQSSMSSEATGEDMPLPSDPLASMGMLPAYESSSEFHSPTTVOVSFP  
 GDGYFLSNVGPPEPFIPTDIPPLSHEDHELTDVATGESHAKNKRRTLELDDPT  
 GSFQRRIRI  
 14331..117588  
 /gene="T28F4.2"  
 join(14331..14423,14659..14790,14924..15073,15122..15276,  
 15394..15523,15569..15721,15766..15840,16226..16358,  
 16407..16579,16662..16802,16866..16940,16988..17089,  
 17463..117588)  
 /note="T28F4.2"  
 /note="Similarity to C.elegans mechanosensory protein 10  
 (SW:ME10\_CHEEL)"  
 /codon\_start=1  
 /protein\_id="CA96693.1"  
 /db\_xref="PID:e1349709"

```
/db_xref="PID:g3880345"
/db_xref="GI:3880345"
/db_xref="SWISS-PROT:Q22851"
/translation="MRGGFVQIFKDESMSTVAVVPVHANANKISRIEWIAFLEF
LGMFAVEYLIAKEFSPATVNTIELERKIFPVVVCNPNKSVYKNSAFSSV
NLTMTYSDATVGTSTDMKGLADNDYDLSRADALVLEANLISDTRKVALVT
YADLIQDSCFAGICSESDPTFTDIPYGAAGNSDALINYSVREGIOGLKMT
VTQIRNGNTDLSPTTLKAGARIGVNSGSSPGDSDNGIDGVCGESEAVSLQNR
AKKPGTCVDPDPSDYKDFITLLETCFNGCQORDIACQCAPPLAAGSTDTAC
QPIKADLCLOTLKGNOTSTPNIDLVECNPCDESTYTPVSLAOPSTSYVA
TSTAGVSCSTNSKESKDCOKWNNNGMIIOVLETLSELYETETAGYVSNYI
NDLGGAGLWGLSVISVEMTGLMVKGAFVGAIKMAPDDEIENDIRIDVED
VKKEIDLEKHGHEGSGSGEDVDIEKKGGEKKK"
complement(18372..20236)
/gene="T28F4.3"
complement(join(18372..18482,18529..18681,18764..19240,
19406..19520,19568..19918,19971..20236))
/gene="T28F4.3"
/notes="Similarity to Human protein-tyrosine phosphatase
epsilon (SW:PTPE_HUMAN)"
/codon_start=1
/protein_id="CA96694.1"
/db_xref="PID:e1349710"
/db_xref="PID:g3880346"
/db_xref="GI:3880346"
/db_xref="SPTREMBL:Q22852"
/translation="MNENDGRNSAKRHNNODESIOPTQDLVDVADAEKSEKSS
NCMIWSQVLLKRAKISNMTREFRNKKEKPVNTISKMDKNRYRIFCGDNRVY
LKERDSDNYIHASWVMPDGFQISTQGLKNTADFMHAYTEKCHVIMLCQYIE
EDQKQCFSTDSERKEGYRVKVEKTEVMKNVKTGLOVKKNCPTHMAYHFE
NSVDVETILKRLMRIGAVOSLOLYIVLVLMLELIDDKIRKRNYSLLAKKA
DVTRKVNIAIMEEKRRQRKREKEKEDEKESPAKREKKEKNGIKTRQKQNS
APSRPNRPEELASKLQGRKRLTKRNKNAEPTAKDKEERSTADKSKNOKQSV
YKPMLESKERKHEKRAKQARKNAKAKKPSVTGNY"
complement(20997..24363)
/gene="T28F4.4"
complement(join(20997..21141,21357..21490,21538..21708,
22192..22363,22407..23017,23241..23352,23410..23591,
23637..24030,24081..24182,24233..24363))
/gene="T28F4.4"
/codon_start=1
/protein_id="CA96695.1"
/db_xref="PID:e1349711"
/db_xref="PID:g3880347"
/db_xref="GI:3880347"
/db_xref="SPTREMBL:Q22853"
/translation="MLNNGFYGLKLANENRREVRILKVLKRNHNVLDROLSSS
GSEHVIAMFEKSDKNROFVIDFSNSICIREOMETERFIEOLLECSIPKRNIAA
FSQTRAGRARLRTSALDVIDYVSSSHIAKATVAFNPHVDTRGITYSENK
FVNTVVRVNNYINNSRCEPSIDIEVQKLSNRNPISDEKINTMRPDIQLEK
DFYSMSYDTPQTSRRKASASLSPYVPSRPSFSSVASASSPAQPSSEGL
DLSLVESENEDEDDKEEKEKVKRKEKKEKKEKKEKKEKKEKKEKKEKKEK
VDCVYVYSTRDPEPTRYRLIRMSSTRHTYIOLDDQFVRLITGCAQPCMLKY
AKSCQCKCSEHREILREFASHVNDNGVAHINSOTKDFVTRTGAIKAKIYK
ERIRAKYPAISEFNALAHYASSPENRREKCKFMDENDEKIAHAKKICENSH
KIKELCEHDMWTFEVEENGECALBTASDDMLHMDENDEKIAHAKKICENSH
FEGWISDFVEKTEGIRTFQVHSESCLEEFERLFLHLLSTCTSTCTVYSAEQATLL
ALSDQPLPTISKLCADDPGLRTILNNGNSLHLPLPALATNAHPTLDTVFLVLR
SSSEVNTKALEICENSTVNTVESEKTFITKC"
26560..26927
/gene="T28F4.5"
join(26560..26742,26793..26927)
/gene="T28F4.5"
/notes="predicted using Genefinder; cDNA EST EMBL:D65871
comes from this gene; cDNA EST EMBL:D69457 comes from this
gene; cDNA EST yk225b4.5 comes from this gene"
/codon_start=1
/protein_id="CA96692.1"
/db_xref="PID:e1349708"
/db_xref="PID:g3880344"
/db_xref="GI:3880344"
/db_xref="SPTREMBL:Q22850"
```

```
/translation="MMDNNTTEEMKVGHPRAKVGGRVYNEKDKNSDNDNANSE
GSDVEAREITIDYLPKMERKSYTEAVKKHEKPMALIDQVHINRANSGVHRFORK
QTH"
complement(27813..28405)
/gene="T28F4.6"
complement(join(27813..28062,28119..28405))
/gene="T28F4.6"
/notes="predicted using Genefinder; cDNA EST yk357e8.5
comes from this gene"
/codon_start=1
/protein_id="CA96696.1"
/db_xref="PID:e1349712"
/db_xref="PID:g3880348"
/db_xref="GI:3880348"
/db_xref="SPTREMBL:Q22848"
/translation="MSANGAIPPKRRSVANDESIONKRSKADTRNEFNIAIK
VEDKVALKAKKILKALPNETKIGSFVSKKLELDVSVFSETSILGTSYHEDH
KSVISILANCCMDKTTSTGMQIKKQSRSLDQAGNIYSPKAVTNOTFFSANKIPI
STQPFKLENGEELORL"
join(28638..29763,29810..30567,272503.1:105..344,
272503.1:397..739,272503.1:796..1322,272503.1:1401..2346,
272503.1:2395..3152,272503.1:3345..3520,
272503.1:3570..3670,272503.1:3718..3789,
272503.1:5050..5460,272503.1:5558..5771,
272503.1:5855..6291,272503.1:6338..6624,
272503.1:6677..6800,272503.1:6847..6966)
/gene="C26C6.1"

Query Match 31.7%; Score 33.6; DB 36; Length 30567;
Best Local Similarity 57.7%; Pred. No. 2.7;
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 1 GGCATTCACATTTAAATATATGTCGATGAGATAGCAACACTGCTCAATAGCC 60
Db 3326 GGCATTAACACACCAAGAGTATGATGATGATGAGAGCAAGCAGTGAAGAGA 3385
OY 61 TAGGCGCTGATTTTGCAGATAATAATAATCATTCATCC 104
Db 3386 TAGGAGAGCTATTCATGAGAAAACAGATTCCTCATTTTCC 3429

RESULT 15
LOCUS STRSCRA 2508 bp DNA BCT 15-JUN-1989
DEFINITION S.mutans enzyme scr-II (scrA) of the phosphoenolpyruvate-dependent
sucrose-6-phosphate hydrolase (scrB; 5' end) genes.
ACCESSION M22711.1 GI:153799
KEYWORDS enzyme scr-II; phosphoenolpyruvate-dependent sucrose
phosphotransferase system; sucrose-6-phosphate hydrolase.
SOURCE S.mutans (strain GS-5) DNA, Clone PMH613.
ORGANISM Streptococcus mutans
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 2508)
AUTHORS Sato,T., Poy,F., Jacobson,G.R. and Kuramitsu,H.K.
TITLE Characterization and sequence analysis of the scrA gene encoding
enzyme Iiscr of the Streptococcus mutans phosphoenolpyruvate-
dependent sucrose phosphotransferase system
J Bacteriol. 171, 263-271 (1989)
J Bacteriol. 171, 263-271 (1989)
MEDLINE 89123027
COMMENT Draft entry and computer-readable sequence for [1] kindly provided
by H.R.Karamitsu, 22-FEB-1989.
FEATURES
location/Qualifiers
1..2508
/organism="Streptococcus mutans"
/db_xref="taxon:1309"
complement(<1..25)
/notes="sucrose-6-phosphate hydrolase (scrB)"
/codon_start=1
/transl_table=1
```

```

/protein_id="AAA26970.1
/db_xref="PID:9552011"
/db_xref="GI:552011"
/translation="NLPONIRY"
CDS
223..2217

```

```

/codon_start=1
/transl_table=11
/protein_id="AAA26971.1"
/db_xref="PID:g153801"
/db_xref="GI:153801"

```

AGDNTNTHETVYNTADYASVEITLASSGTYAABDSLEJENK -  
 DNGAAYEVLAPAGEVELTIVNDPESSAMGGVAIKRSGNYTVAPYNGTOIAF  
 KAOTYEVLAPAGEVELTIVNDPESSAMGGVAIKRSGNYTVAPYNGTOIAF  
 AGDNTNTHETVYNTADYASVEITLASSGTYAABDSLEJENK -

```
misc_signal      2227.. 2252 transcription termination signal; putative"
BASE COUNT      740 a 401 c 513 g 854 t
ORIGIN           3 bp upstream of EcoRV site.
```

Query Match	30.68;	Score 32.4;	DB 1;	length 2508;
Best Local Similarity	62.28;	Pred. No. 6.1;		
Matches	51;	Conservative	0;	Mismatches 31;
			Indels	0;
			Gaps	0;

QY 14 TAAATATTAGCTGGAGCTGGATTAAGGAACTGCACGCTGTCAATAAGCTTAGGGCTGAATTT 73  
| | | | | | | | | | | | | | | |  
Db 1050 TGACAATAAAGCTGTAAAGTAAAGTCACACTGCGTAGGGCAATCTCTTATGSAAAAATTT 991

QY	74	TTGTCAGATAAATAAATAAAT	95
Db	990	TTCTAAGATAGATAAAAGCCAT	969

Search completed: August 5, 1999, 23:16:24  
Job time: 7867 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:16:24 ; Search time 3264.21 Seconds

(without alignments)  
2740.697 Million cell updates/sec

Title: US-09-049-696-18

Perfect score: 2813

Sequence: 1 GAAATCAGCAGGAGATGTAC.....AATAATATCATTCCTTA 2813

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenBank:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_htg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_lm:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_htg1:\*  
35: gb\_htg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2807.2	99.8	3311	11	AF039400	Homo sapi

Result No.	Score	Match	Length	DB	ID	Description
2	2790.4	99.2	2836	42	AF127036	AF127036 Homo sapi
3	1743	62.0	2937	12	AB017156	AB017156 Mus muscu
4	933.4	33.2	2984	3	BT036445	BT036445 Bos tauru
5	900.2	32.0	3317	3	AF001261	AF001261 Bos tauru
6	863	30.7	3022	12	AF047838	AF047838 Mus muscu
7	863	30.7	3471	12	AF052746	AF052746 Mus muscu
8	831	29.5	3415	42	AF043976	AF043976 Homo sapi
9	806	28.7	2820	3	AF001263	AF001263 Bos tauru
10	800.6	28.5	3288	3	AF001262	AF001262 Bos tauru
11	780.8	27.8	878	5	195746	195746 Sequence 8
12	550.6	19.6	3604	9	AB026833	AB026833 Homo sapi
13	435.4	15.5	35278	11	AF039401	AF039401 Homo sapi
14	381	13.5	1265	3	AF001264	AF001264 Bos tauru
15	52	1.8	7218	5	166494	166494 Sequence 14
16	41.6	1.5	1231	1	MC082	MC082 Parametrium
17	40.6	1.4	330742	17	PM042580	PM042580 Caenorhab
18	40.6	1.4	99083	35	CEY37A1B	CEY37A1B Plasmodiu
19	40.2	1.4	114226	35	AC004710	AC004710 Plasmodiu
20	40.2	1.4	258033	35	AC005507	AC005507 Plasmodiu
21	40.2	1.4	11534	36	PEAARP1PR	PEAARP1PR Plasmodiu
22	40	1.4	179923	35	AC007246	AC007246 Homo sapi
23	39	1.4	2937	2	AF036004	AF036004 Pasteurel
24	39	1.4	16724	2	AF067175	AF067175 Pasteurel
25	39	1.4	164399	36	PFMAL3P6	PFMAL3P6 Plasmodium
26	39	1.4	164399	36	PFMAL3P6	PFMAL3P6 Plasmodium
27	38.8	1.4	40450	36	CEC54D10	CEC54D10 Plasmodium
28	38.6	1.4	29324	1	BS16829KB	BS16829KB Bacillus
29	38.6	1.4	3215	1	BSCEVEGE	BSCEVEGE Bacillus
30	38.6	1.4	3215	1	BSU05345	BSU05345 Bacillus su
31	38.6	1.4	208230	7	BSUB0008	BSUB0008 Bacillus su
32	38.6	1.4	94695	1	ATF23E13	ATF23E13 Arabidops
33	38.6	1.4	8700	8	AF068574	AF068574 Arabidops
34	38.6	1.4	189165	35	AC007237	AC007237 Homo sapi
35	38.4	1.4	1691	4	CHRC1P2M	CHRC1P2M Chicken typ
36	38.4	1.4	1248	4	CHRCOLA29	CHRCOLA29 Chicken ald
37	38.4	1.4	1683	4	GCCO10	GCCO10 Chicken pro
38	38.4	1.4	1010	4	GCCOLA2C	GCCOLA2C Chicken pro
39	38.4	1.4	122459	35	AC006944	AC006944 Mus muscu
40	38.2	1.4	79355	7	AB012247	AB012247 Arabidops
41	38.2	1.4	184949	11	AC005029	AC005029 Homo sapi
42	38.2	1.4	167870	35	AC007463	AC007463 Homo sapi
43	38	1.3	160378	2	HS93H18	HS93H18 Human DNA s
44	37.8	1.3	19436	9	U67522	U67522 Methanococ
45	37.8	1.3	92558	9	HS341E18	HS341E18 Human DNA

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	MRNA	PRI	15-DEC-1998
LOCUS	AF039400					
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hccal1) mRNA, complete cds.					
ACCESSION	AF039400					
NID	94009457					
VERSION	AF039400.1	GI:4009457				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
AUTHORS	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.					
TITLE	Genomic cloning, molecular characterization, and functional analysis of human CACAL1, the first human member of the family of Ca2+-activated Cl- channel proteins					
JOURNAL	Genomics 54 (2), 200-214 (1998)					
MEDLINE	99047526					
REFERENCE	2 (bases 1 to 3311)					
AUTHORS	Gruber,A.D., Elble,R. and Pauli,B.U.					
TITLE	Direct Submission					

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers  
source 1. 3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. 3311  
/gene="hCLCA1"  
352. 3096  
/gene="hCLCA1"  
/note="Transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="pid:94009458"  
/db\_xref="gi:4009458"  
/translation="WGPKSVFTLLILHLEGLASNSLIQLNNNGEGIVVAIDPNP  
EDELIIQIKDMVTOASLYLEFATGRKRFYFNKVALILEPTWKADYVRPLEYKMA  
DVLVAESTPGNDEBYTEMGNCKEGRIRHLPFIFIGKLAEGPGKAFVHEMAH  
LRMGVEDEYNDEKFTYLSNGRIQAVRSGAGITGVVKKCGGSGCYTRGFENKFTGL  
YKGEPEYLOSROTEKASIMRQHYDSIVERCTEQNHKKEAPNKNOKRSTMEVI  
RDESEFKTTPPTDPPNPFTSLDIOGRIVCLVDKSGSMATGRRLRNDAGOLE  
LQTVELGSMGVAVTFDSAHHVQSELDIQNSGSDROTLLKRLPAAGSGTSLCSGRSA  
FTVIRKKYPTDSEIYLVDGEDNTISGFNEVKSGAIHTVALGPSAADELELSK  
MTGLOTVASDOVONNGLIDAFGLSNGAVSIOLESGLTLNKAAYLQIGIKAVGWKY  
DSYVKDPLFIITWTPOILLMPDSGOKGFOVDNKTKAAYLQIGIKAVGWKY  
SLOASQTLTLTSTRSANNATLPITVSKTNKDSKRPSPVAVYANIRGASGTLRA  
SVTALESVNGKTVLELLDNGAGADARKDGVYRRTTYTDNRYSVKRYRAGVN  
AARRVIVQOSGALYIPGMWENDELQMPREPKNDVQHVQVCFSTVSGGSVAS  
DWNAPILDEPPQITDLKAEIHGSLINLTWAPGDYDHGTAHKYIIRISILD  
LRDNESLQVWTLALIKRANSEVFLFEPENITFEFGTDLFIHQADVYDLKSEI  
SNIRVSLFIPQPTPEPSPDETSAPCPNHIHINSTIPGHIHKIMMKIGELQLSIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 99.8%; Score 2807.2; DB 11; Length 3311;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2809; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAATCAGAGGAGATGTACAGCAATGGGGCCATTTAAGATTCTGTTCATCTTGATT 60  
DB 328 GGAATCAGAGGAGATGTACAGCAATGGGGCCATTTAAGATTCTGTTCATCTTGATT 387  
QY 61 CTTACCTCTAGAGAGGGCCCTGATTAATTCATCTCATAGCTGAAACAACATGGCTAT 120  
DB 388 CTTACCTCTAGAGAGGGCCCTGATTAATTCATCTCATAGCTGAAACAACATGGCTAT 447  
QY 121 GAAGGATGTGCTTGCATTCGACCCCAATGTGCCAGAGATGAACACATCTCAACAA 180  
DB 448 GAAGGATGTGCTTGCATTCGACCCCAATGTGCCAGAGATGAACACATCTCAACAA 507  
QY 181 ATTAAGACATGGTGACCAAGGACATCTGTATCTGTTGAAGCTACAGAAAGCATTT 240  
DB 508 ATTAAGACATGGTGACCAAGGACATCTGTATCTGTTGAAGCTACAGAAAGCATTT 567  
QY 241 TATTTCAAAAATGTTGCCATTTTGAATTCCTGAAACATGGAAGCAAAAGGCTGACTATG 300  
DB 568 TATTTCAAAAATGTTGCCATTTTGAATTCCTGAAACATGGAAGCAAAAGGCTGACTATG 627  
QY 301 AGACCAAACTTGAGACCTACAAAATGCTGATGTCTGGTTCAGTCTACTCTCTCA 360  
DB 628 AGACCAAACTTGAGACCTACAAAATGCTGATGTCTGGTTCAGTCTACTCTCTCA 687  
QY 361 GGTATATGATGAACCTTACACTGAGCAGATGGGCACTGTGAGAGAGAAAGGTGAAGATC 420  
DB 688 GGTATATGATGAACCTTACACTGAGCAGATGGGCACTGTGAGAGAGAAAGGTGAAGATC 747  
QY 421 CACCTACCTCGATTTCAATGTGACAGAAAAAATTAGCTGAATATGACACCAAGGTAG 480

DB 748 CACCTACCTCGATTTATTTGACAGAAAAAAGTAGCTGATATGAGCCACAAAGGTAG 807  
QY 481 GCATTTGTCAATGAGTGGGCTCATCTACGATGGGAGATTTTGACGATACAAATATGAT 540  
DB 808 GCATTTGTCAATGAGTGGGCTCATCTACGATGGGAGATTTTGACGATACAAATATGAT 867  
QY 541 GAGAAATTTCTACTTATCCATGGAAGAAATACAGCAGTAAAGATGTTGACGAGATTTACT 600  
DB 868 GAGAAATTTCTACTTATCCATGGAAGAAATACAGCAGTAAAGATGTTGACGAGATTTACT 927  
QY 601 GGTACAAATGTAGTAAAGAGTGCAGGAGGAGGAGGAGCTTTACACCAAGATGACATTC 660  
DB 928 GGTACAAATGTAGTAAAGAGTGCAGGAGGAGGAGGAGGAGCTTTACACCAAGATGACATTC 987  
QY 661 AATAAGTAAAGAGCTATGAAAAAGATGTGAGTTGTTCTCCAAATCCCGCCAGAG 720  
DB 988 AATAAGTAAAGAGCTATGAAAAAGATGTGAGTTGTTCTCCAAATCCCGCCAGAG 1047  
QY 721 GAGAGGCTTCTATATATGTTTGACACACATGTTGATTTCTATAGTTGAATTTCTGACAGA 780  
DB 1048 GAGAGGCTTCTATATATGTTTGACACACATGTTGATTTCTATAGTTGAATTTCTGACAGA 1107  
QY 781 CAAACCAACCAAGAGCTCCAAACAGCAAGCAAAATCAAAAATGCAATCTCCGAAGACA 840  
DB 1108 CAAACCAACCAAGAGCTCCAAACAGCAAGCAAAATCAAAAATGCAATCTCCGAAGACA 1167  
QY 841 TGGAGATGATCCGATGATCTGAGACCTTTAAGAAACCACTCTATGACACACAGCA 900  
DB 1168 TGGAGATGATCCGATGATCTGAGACCTTTAAGAAACCACTCTATGACACACAGCA 1227  
QY 901 CCAATTCACACTTCTCATTTGTCGAGATTTGGACAAATATGTTGTTAGTCTCTTAC 960  
DB 1228 CCAATTCACACTTCTCATTTGTCGAGATTTGGACAAATATGTTGTTAGTCTCTTAC 1287  
QY 961 AATCTGGAAGATGGCACTGATGATTAACCGCTCAATTCAGATGAATCAAGCAGGCGAGCT 1020  
DB 1288 AATCTGGAAGATGGCACTGATGATTAACCGCTCAATTCAGATGAATCAAGCAGGCGAGCT 1347  
QY 1021 TTCTGCTGACAGACATTTGAGCTGGGCTCTGGGTTGGATGATGACATTTGACAGTCT 1080  
DB 1348 TTCTGCTGACAGACATTTGAGCTGGGCTCTGGGTTGGATGATGACATTTGACAGTCT 1407  
QY 1081 GCCCATGTACAAAGTGAATCTATACAGATTAACAGTGGCAGTGAAGGACACACTCGCC 1140  
DB 1408 GCCCATGTACAAAGTGAATCTATACAGATTAACAGTGGCAGTGAAGGACACACTCGCC 1467  
QY 1141 AAAAGATTACCTGACAGACCTTCAGAGAGGAGCTCATCTGACGCGGCTTGATCGGCA 1200  
DB 1468 AAAAGATTACCTGACAGACCTTCAGAGAGGAGCTCATCTGACGCGGCTTGATCGGCA 1527  
QY 1201 TTTACTGTGATTTAGAGAAATATCAACTGATGATCTGAATTTGTCTGCTGACGAT 1260  
DB 1528 TTTACTGTGATTTAGAGAAATATCAACTGATGATCTGAATTTGTCTGCTGACGAT 1587  
QY 1261 GGGGAAGCAACATTAAGTGGGTCTTTAAGAGGTCAACAAAGGTGGCCATCATC 1320  
DB 1588 GGGGAAGCAACATTAAGTGGGTCTTTAAGAGGTCAACAAAGGTGGCCATCATC 1647  
QY 1321 CACACAGTGTGTTGGGCGCTCTGACGCTCAAGAACTAAGAGAGCTGTCCAAATGACA 1380  
DB 1648 CACACAGTGTGTTGGGCGCTCTGACGCTCAAGAACTAAGAGAGCTGTCCAAATGACA 1707  
QY 1381 GGAGGTTTACAGACATATGCTTACAGATCAAGTTCAGAAACATGGCTATGATCTTTT 1440  
DB 1708 GGAGGTTTACAGACATATGCTTACAGATCAAGTTCAGAAACATGGCTATGATCTTTT 1767  
QY 1441 GGGGCGCTTCACTACAGAAATGAGAGTGTCTGAGCGCTCATCAGCTTGAGAGTAAG 1500  
DB 1768 GGGGCGCTTCACTACAGAAATGAGAGTGTCTGAGCGCTCATCAGCTTGAGAGTAAG 1827  
QY 1501 GGATTAACTTCAGAAAGCAGTGAATGGACAGTATGTCGACAGCACCGTG 1560

Db 1828 GGATTAAACCTCCAGAACGCCAGTGGATGAATGACACAGTGTGCTGGACAGCACCCTG 1887  
 Qy 1561 GGAAGACACTTTGTTCTTATCACCCTGACACAGCAGCTCCCAAAATCCTTCTG 1620  
 Db 1888 GGAAAGACACTTTGTTCTTATCACCCTGACACAGCAGCTCCCAAAATCCTTCTG 1947  
 Qy 1621 GATCCAGTGGACAGAACAGAGTGGCTTTGTAGTGGACAAAAACCAAAAGGCTTAC 1680  
 Db 1948 GATCCAGTGGACAGAACAGAGTGGCTTTGTAGTGGACAAAAACCAAAAGGCTTAC 2007  
 Qy 1681 CTCCAATCCAGGACATCTTAAGTTGGACATGGAAATACAGTCTGCAAGCACTCA 1740  
 Db 2008 CTCCAATCCAGGACATCTTAAGTTGGACATGGAAATACAGTCTGCAAGCACTCA 2067  
 Qy 1741 CAACCTTGACCTGACCTGACGCTCCGCTGCTCAATGCTACCTGCTCCCAATTACA 1800  
 Db 2068 CAACCTTGACCTGACCTGACGCTCCGCTGCTCAATGCTACCTGCTCCCAATTACA 2127  
 Qy 1801 GTGACTTCCAAAACGAACAGACAGACACCAATTCGCCAGCTCTGGTAGTTATGCA 1860  
 Db 2128 GTGACTTCCAAAACGAACAGACAGACACCAATTCGCCAGCTCTGGTAGTTATGCA 2187  
 Qy 1861 AATATTCGCAAGAGACCTCCCAATTCAGGCGCAGTGTCAAGCCTGATTTGATCA 1920  
 Db 2188 AATATTCGCAAGAGACCTCCCAATTCAGGCGCAGTGTCAAGCCTGATTTGATCA 2247  
 Qy 1921 GTGATGGAACAAACAGTACTCTTGAACACTGATGATGAGACAGTGTCTGATCT 1980  
 Db 2248 GTGATGGAACAAACAGTACTCTTGAACACTGATGATGAGACAGTGTCTGATCT 2307  
 Qy 1981 AAGATGAGAGTGTCTACTCAAGTATTTCAACCTTATGACACGAATGGTAGATCACT 2040  
 Db 2308 AAGATGAGAGTGTCTACTCAAGTATTTCAACCTTATGACACGAATGGTAGATCACT 2367  
 Qy 2041 GTAAAGTGGGCTCTGGGAGAGTTAAAGCAGCAGAGAGTGTATACCCAGCAG 2100  
 Db 2368 GTAAAGTGGGCTCTGGGAGAGTTAAAGCAGCAGAGAGTGTATACCCAGCAG 2427  
 Qy 2101 AGTGGACACTGTACATACCTGGCTGGATGAGATGATGAATGAATGAATGCCACA 2160  
 Db 2428 AGTGGACACTGTACATACCTGGCTGGATGAGATGATGAATGAATGAATGCCACA 2487  
 Qy 2161 AACCCGAATTAATGAATGATGTTCAACCAAGTGTGTTCAAGCAATATCC 2220  
 Db 2488 AACCCGAATTAATGAATGATGTTCAACCAAGTGTGTTCAAGCAATATCC 2547  
 Qy 2221 TGGGAGGCTCATTTGTGCTTCTGATGTCCTCAATGCTCCATACCTGATCTTCCA 2280  
 Db 2548 TGGGAGGCTCATTTGTGCTTCTGATGTCCTCAATGCTCCATACCTGATCTTCCA 2607  
 Qy 2281 CCTGGCAATATCACGACCTGAAGGGGGAATTCACGGGGGAGTCTCATTAATCTGACT 2340  
 Db 2608 CCTGGCAATATCACGACCTGAAGGGGGAATTCACGGGGGAGTCTCATTAATCTGACT 2667  
 Qy 2341 TGGACAGCTCTTGGGATGATTTATGACCATGGAAGCTTCACAGATATCATTCGAATA 2400  
 Db 2668 TGGACAGCTCTTGGGATGATTTATGACCATGGAAGCTTCACAGATATCATTCGAATA 2727  
 Qy 2401 AGACAAAGTATCTTGTATCTCAGAGACAGATGAATGAATCTTCAAGGAATCTACT 2460  
 Db 2728 AGACAAAGTATCTTGTATCTCAGAGACAGATGAATGAATCTTCAAGGAATCTACT 2787  
 Qy 2461 GCTCTCATCCCAAGAGAACCACTCTGAAGAGTCTTTTGTGTTAAACGAGAAACAT 2520  
 Db 2788 GCTCTCATCCCAAGAGAACCACTCTGAAGAGTCTTTTGTGTTAAACGAGAAACAT 2847  
 Qy 2521 ACTTTGGAATGAGCAGACATCTTTTCATTTGCTATTCAGGCTTGTATAAGTCGATCG 2580  
 Db 2848 ACTTTGGAATGAGCAGACATCTTTTCATTTGCTATTCAGGCTTGTATAAGTCGATCG 2907  
 Qy 2581 AATTCGAATATCCAAAGTTCAGAGATATCTTTTATTCTCCACAGACTCCGCCA 2640  
 Db 2908 AATTCGAATATCCAAAGTTCAGAGATATCTTTTATTCTCCACAGACTCCGCCA 2967

Qy 2641 GAGACACCTAGTCTGATGAACAGTCTGCTCTTCTTCAATATTCATATCAACAGCACC 2700  
 Db 2968 GAGACACCTAGTCTGATGAACAGTCTGCTCTTCTTCAATATTCATATCAACAGCACC 3027  
 Qy 2701 ATTCCCTGACATTCATTTAAATATGAGAGTGGATGAGAACTGAGAGCTGCA 2760  
 Db 3028 ATTCCCTGACATTCATTTAAATATGAGAGTGGATGAGAACTGAGAGCTGCA 3087  
 Qy 2761 ATAGCCTAGGCTGATTTTGTGATGAATGAATGAATGAATGAATGAATGAATGAAT 2812  
 Db 3088 ATAGCCTAGGCTGATTTTGTGATGAATGAATGAATGAATGAATGAATGAATGAAT 3139

RESULT 2  
 AF127036  
 LOCUS Homo sapiens  
 DEFINITION Homo sapiens calcium-activated chloride channel protein 1 (CACCL1)  
 ACCESSION AF127036  
 NID 94585468  
 VERSION AF127036.1 GI:4585468  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 2826)  
 Aguel, M., Vermaat, T., and Culouscou, J.-M.  
 Cloning of three human homologs of bovine epithelial chloride channel  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2826)  
 Aguel, M., and Culouscou, J.-M.  
 Direct Submission  
 TITLE Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
 Carrières, Neuilly-Malmaison 92500, France  
 FEATURES  
 source  
 1..2826  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="small intestine; colon"  
 1..2826  
 /gene="CACCL1"  
 5..2749  
 /note="bovine epithelial chloride channel homolog"  
 /product="calcium-activated chloride channel protein 1"  
 /codon\_start=1  
 /protein\_id="AA025487.1"  
 /db\_xref="PID:94585469"  
 /translation="MGPKSSVFILHLLEGALSNSLIQNNNGEIVAIIDPNV  
 EDLILQIKMTQASLYLEFATGKRYEFNVALLPEMTKADYRPLRYKKA  
 DIVYAEIPEPNDPEPTROMKNGCEKGRHILTFPLAGKLAIEGPGRAFVEMAH  
 LRMVFEIDNDEKFTLSNGRIQARVCAIGTVAKKCGGSGCTIRKFNKVTGL  
 YKGECEFLQROTEKASIMFAQHDVSVECTQNNHKNPNQONCNRSTWEVI  
 RDESEFKTTPMTQPPNPFTSLQIGRIQVLDKSGSMATGRNLRNLQAGLFL  
 LQTEVLSGVMATFSDSAHVOSELIQNSGSDRLTAKRLPAASGCTSGISGLRA  
 FTVIRKRYPTGSEIVILLDGEDMTISCFNEVROSGAIIHTVALGPEAAQLEELSK  
 MTGGLQYASQVQNNCLIDAFGLSSNGNAVSRISDLESGLTQNSQNMNGVIV  
 DSTVGEKDLFLITVTPQILLMDPSQKGGGVVDNLTMAVLIQIAGVWKI  
 SLQASQDTLLITVSRASNATLPITYTSKNTKTSKPSPLVYVANTROGASPLRA  
 SVTALIESVNGKTVTLLELDNGADATKDGVSRYETTYDINGRSKVRALGVN  
 AARRRVIPQOGALYIPGMIENDEIQNNPPEINKDVOYKQVCFSSSGSEFVAS  
 DVNAPRPDLEPGQITDLKAEIHGGSILNTLTAPOGDYHGTAAHKYILIRISIID  
 LROKFNESLQVNTALIPKENSSEVLEFKENITPENGTDLFLAIQAVDVPDLAKSEI  
 SNIAKVSLEIPQIPPEPSPDETSAPEPNHINSTIIPGIIHLIMKRWIGELDLSTA

BASE COUNT 875 a 623 c 632 g 696 t  
 ORIGIN

Query Match 99.28; Score 2790.4; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

21 AGCAATGGGGCCATTAAAGAGTCTGTTCTCATCTTGAATCTTCACTCTAGAGGGG 80  
1 AGCAATGGGGCCATTAAAGAGTCTGTTCTCATCTTGAATCTTCACTCTAGAGGGG 60  
81 CCTGATTAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 140  
61 CCTGATTAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 120  
141 CGACCCCAATGTCCAGAGATGAAACACTCATTCACTCACTCACTCACTCACTCACT 200  
121 CGACCCCAATGTCCAGAGATGAAACACTCATTCACTCACTCACTCACTCACTCACT 180  
201 GGCATCTCTGATCTGTTGAGAGTACAGGAAGGATTTTATTTCAAAAATGTTCCAT 260  
181 GGCATCTCTGATCTGTTGAGAGTACAGGAAGGATTTTATTTCAAAAATGTTCCAT 240  
261 TTTGATTCCTGAACATGGAAGCAAAAGGCTGACTATGTGAGACCAAACTTGAGACTA 320  
241 TTTGATTCCTGAACATGGAAGCAAAAGGCTGACTATGTGAGACCAAACTTGAGACTA 300  
321 CAAAATGCTGATGTTCTGTTCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 380  
301 CAAAATGCTGATGTTCTGTTCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 360  
381 TGAAGAGTGGGCACTGTGAGAGAGGCTGAAAGGATTCACCTGACTCTGATTTTCA 440  
361 TGAAGAGTGGGCACTGTGAGAGAGGCTGAAAGGATTCACCTGACTCTGATTTTCA 420  
441 TGCAGAAAAAAGTTAGCTAATATGAGCAAGGAGGCTTTGTCATGAGTGGGC 500  
421 TGCAGAAAAAAGTTAGCTAATATGAGCAAGGAGGCTTTGTCATGAGTGGGC 480  
501 TCACTACGATGGGAGTATTTGACAGATCAATATGATGAGAAATTTCTACTATCCAA 560  
481 TCACTACGATGGGAGTATTTGACAGATCAATATGATGAGAAATTTCTACTATCCAA 540  
561 TGAAGAGTACAGCACTAGATGTTCAAGCAAGTATCTGATCAATATGATGAGAA 620  
541 TGAAGAGTACAGCACTAGATGTTCAAGCAAGTATCTGATCAATATGATGAGAA 600  
621 GTGTCAGGAGGAGCTGTACACCAAAAGATGCAATTCATTAAGTAAAGAGACTCTA 680  
601 GTGTCAGGAGGAGCTGTACACCAAAAGATGCAATTCATTAAGTAAAGAGACTCTA 660  
681 TGAAGAGATGATGATTTGTTCTCCATCCCGCAGAGGAGGAGGCTTCTATATGTT 740  
661 TGAAGAGATGATGATTTGTTCTCCATCCCGCAGAGGAGGAGGCTTCTATATGTT 720  
741 TGCACAATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 800  
721 TGCACAATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
801 TCCAAACAGCAAAATCAAAAATGCAATCTCGAAGCACTGGAAGATGATCGATTC 860  
781 TCCAAACAGCAAAATCAAAAATGCAATCTCGAAGCACTGGAAGATGATCGATTC 840  
861 TGAAGACTTTAAGAAACCACTCTATGCAACAGCAGCAGCAATCCACCTCTCAT 920  
841 TGAAGACTTTAAGAAACCACTCTATGCAACAGCAGCAGCAATCCACCTCTCAT 900  
921 GCTGAGATTTGAGAAACCACTCTATGCAACAGCAGCAGCAATCCACCTCTCAT 980  
901 GCTGAGATTTGAGAAACCACTCTATGCAACAGCAGCAGCAATCCACCTCTCAT 960  
981 TGGTAACCGCTCAATCGATGATCAAGAGGCGCAGCTTTCTCTGTCAGACAGTTGA 1040  
961 TGGTAACCGCTCAATCGATGATCAAGAGGCGCAGCTTTCTCTGTCAGACAGTTGA 1020  
1041 GCTGGAGCTCTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1100

1021 GCTGGAGCTCTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
1101 CATACAGTAAACAGTGGCAGTGCAGAGGACACACTGCGCAAAAGATTTACTGACAG 1160  
1081 CATACAGTAAACAGTGGCAGTGCAGAGGACACACTGCGCAAAAGATTTACTGACAG 1140  
1161 TTCAGAGGAGAGTCCATCTGACCGGCTTCGATCGGCATTTTCTGATTTGAGAA 1220  
1141 TTCAGAGGAGAGTCCATCTGACCGGCTTCGATCGGCATTTTCTGATTTGAGAA 1200  
1221 AATCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1280  
1201 AATCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
1281 TGGGCTTTAAGAGAGTCAAAACAAAGTGGCCATCATCCACACAGTGGCTTGGGGCC 1340  
1261 TGGGCTTTAAGAGAGTCAAAACAAAGTGGCCATCATCCACACAGTGGCTTGGGGCC 1320  
1341 CTCTGAGCTCAAGACTAGAGAGCTGCTCCAAATGACAGAGGTTTACAGATATGC 1400  
1321 CTCTGAGCTCAAGACTAGAGAGCTGCTCCAAATGACAGAGGTTTACAGATATGC 1380  
1401 TTCAGATCAAGTTCAGAAACATGAGCTCATTTGATGCTTTTGGGGCCCTTTATCAGAAA 1460  
1381 TTCAGATCAAGTTCAGAAACATGAGCTCATTTGATGCTTTTGGGGCCCTTTATCAGAAA 1440  
1461 TGAAGCTCTCTCAGGCTCCATCCAGCTGAGATGAGATGAGATGAGATGAGATGAG 1520  
1441 TGAAGCTCTCTCAGGCTCCATCCAGCTGAGATGAGATGAGATGAGATGAGATGAG 1500  
1521 CCAAGTATGATGAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1580  
1501 CCAAGTATGATGAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
1581 TATCACTGAGCAAGCAGCTCCCAAAATCCTCTCTGAGATCCAGTGCAGAGCA 1640  
1561 TATCACTGAGCAAGCAGCTCCCAAAATCCTCTCTGAGATCCAGTGCAGAGCA 1620  
1641 AGGTGGCTTTGATGAGCAAAACACCAAAATGAGCTTACCTCAATCCAGGACTTGC 1700  
1621 AGGTGGCTTTGATGAGCAAAACACCAAAATGAGCTTACCTCAATCCAGGACTTGC 1680  
1701 TAAAGTTGGCACTTGAATATCACTGTCAGCAAGTATGATGATGATGATGATGATGAT 1760  
1681 TAAAGTTGGCACTTGAATATCACTGTCAGCAAGTATGATGATGATGATGATGATGAT 1740  
1761 CACGTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1820  
1741 CACGTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
1821 GAGACACAGCAAAATCCCGACCTCTGATGATGATGATGATGATGATGATGATGATGAT 1880  
1801 GAGACACAGCAAAATCCCGACCTCTGATGATGATGATGATGATGATGATGATGATGAT 1860  
1881 CCCAATTTCAAGGAGCAGTGTACAGCCTGATGATGATGATGATGATGATGATGATGAT 1940  
1861 CCCAATTTCAAGGAGCAGTGTACAGCCTGATGATGATGATGATGATGATGATGATGAT 1920  
1941 CTTGGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2000  
1921 CTTGGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
2001 AAGTATTTCAACACTTATGACAGCAAGTATGATGATGATGATGATGATGATGATGAT 2060  
1981 AAGTATTTCAACACTTATGACAGCAAGTATGATGATGATGATGATGATGATGATGAT 2040  
2061 AAGGATTTAAGCAGCAGCAGAGAGATGATACCCAGCAGAGTGCATGATATCC 2120  
2041 AAGGATTTAAGCAGCAGCAGAGAGATGATACCCAGCAGAGTGCATGATATCC 2100  
2121 TGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2180



```

Db 2101 TGGCTGGATTGAGATGATGAATTAATACATGGAATCCACCAGACTGAAATTAATAGGA 2160
Oy 2181 TGATGTTCAACACAGCAAGTGTGTTTCAGCAGACATCCTCGGAGGCTCATTTGTGGC 2240
Db 2161 TGATGTTCAACACAGCAAGTGTGTTTCAGCAGACATCCTCGGAGGCTCATTTGTGGC 2220
Oy 2241 TTTCTGATGCCCAATGCTGCCATACCTGATGTTCTTCTTCCACCTGGCCAAATACCGACT 2300
Db 2221 TTTCTGATGCCCAATGCTGCCATACCTGATGTTCTTCTTCCACCTGGCCAAATACCGACT 2280
Oy 2301 GAAGGGGAATTCACGGGGGAGCTCATTAATTCGACTTGAGACAGCTCTGGGATGA 2360
Db 2281 GAAGGGGAATTCACGGGGGAGCTCATTAATTCGACTTGAGACAGCTCTGGGATGA 2340
Oy 2361 TTATGACCATGGAACAGCTCACAAGATATATCATTCGAAATAGTACAAGTATTCCTGATCT 2420
Db 2341 TTATGACCATGGAACAGCTCACAAGATATATCATTCGAAATAGTACAAGTATTCCTGATCT 2400
Oy 2421 CAGAGCAAGTTCAATGAATCTCTTCAAGTATCTACTGCTCTCATCCCAAGGAGC 2480
Db 2401 CAGAGCAAGTTCAATGAATCTCTTCAAGTATCTACTGCTCTCATCCCAAGGAGC 2460
Oy 2481 CAACCTGAGGAAGTCTTTTGTGTTAAACAGAAAACATTACTTTGAAATGGCAGAGA 2540
Db 2461 CAACCTGAGGAAGTCTTTTGTGTTAAACAGAAAACATTACTTTGAAATGGCAGAGA 2520
Oy 2541 TCTTTTCATTGCTATTCAGGCTGTGTAAGTGCATCTGAAAATCAGAAATATCAACAT 2600
Db 2521 TCTTTTCATTGCTATTCAGGCTGTGTAAGTGCATCTGAAAATCAGAAATATCAACAT 2580
Oy 2601 TGCACAGATATCTTTTATTATTCCTCAGACAGATCCGCCAGACACACTAGTCTGATGA 2660
Db 2581 TGCACAGATATCTTTTATTATTCCTCAGACAGATCCGCCAGACACACTAGTCTGATGA 2640
Oy 2661 AACGTCGCTGCTGCTGCTCAATATATCATATCAACAGACACACTTCGGCATTCACATTTT 2720
Db 2641 AACGTCGCTGCTGCTGCTCAATATATCATATCAACAGACACACTTCGGCATTCACATTTT 2700
Oy 2721 AAAAATTAATGAGAGTGTAGAGAGCAAGCTGCTCAATAGCCTAGGGCTGAATTTT 2780
Db 2701 AAAAATTAATGAGAGTGTAGAGAGCAAGCTGCTCAATAGCCTAGGGCTGAATTTT 2760
Oy 2781 TGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2840
Db 2761 TGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2820

RESULT 3
AB017156 2937 bp mRNA ROD 04-MAR-1999
LOCUS AB017156 Mus musculus gob-5 mRNA, complete cds.
DEFINITION AB017156
ACCESSION 93721911
NID 93721911
VERSION AB017156.1 GI:3721911
KEYWORDS GOB-5.
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites) Tanigawa,Y. and Hirohashi,S.
AUTHORS Komiya,T., Tanigawa,Y. and Hirohashi,S.
TITLE Cloning and identification of the gene gob-5, which is expressed in
intestinal goblet cells in mice
JOURNAL Biochem. Res. Commun. 255 (2), 347-351 (1999)
MEDLINE 99160866
REFERENCE 2 (bases 1 to 2937)
AUTHORS Komiya,T., Tanigawa,Y. and Hirohashi,S.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/Genbank databases. Tohtu
Komiyu, ERATO, JST, Hirohashi Cell Configuration Project; 5-9-9,
Tokodai, Tsukuba, Ibaraki 300-2635, Japan
(E-mail:tkom@nccp.jst.go.jp, Tel:81-298-47-7563,
Fax:81-298-47-5226)

```

```

FEATURES
    source          1..2937
                    Location/Qualifiers
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /cell_type="goblet cell"
                    /dev_stage="adult"
                    /tissue_type="intestine"
    gene            15..2756
                    /gene="gob-5"
    CDS              15..2756
                    /gene="gob-5"
                    /codon_start=1
                    /protein_id="BA33743.1"
                    /db_xref="PID:01034712"
                    /db_xref="PID:93721912"
                    /db_xref="GI:3721912"
                    /translation="MESLSKSPVFLILHLLEGVLESLLIQNNNGEIVIAIDHDPV
                    EDEALIOHKMVTQASPYLFEATGKRRFFRNVAIILIPESMKAREYTRPMEKKA
                    DIVSTSTPLINDPEYTHIGACGEGKRIHLTDPLAGKRLTOYGGODRPFHEHMAH
                    FRGVFNENDEKFTYLSKRPQAVRCSAATGKNQVRCGGSCITNGKCYIDRYTG
                    LYKDCVFPDPHONKASIMFNQINSVEFEKRNQENPQDNCNMLRTWEY
                    IQESDFKQTPMTPAQPPFSLQIQRIQVCLVLDKRGSLNDBDRMNOASRLF
                    LLQYEOGSMVGMTFDSAAYVSEIKQNSGADRLDLIKHLPTVSAGSTISGLRT
                    AFTVTKRYPIDPQIIVLTGDEDMTISCFDLYVQSGAIIHTVILGPAARKEIOLS
                    KMTGGLQYSSDQVQNNGLVDAPFALSSGNAIAQHSIQLESRCVNLQNNQMSVY
                    VDSVQKDTLELITWTHPTTFLIDPVGEDQNGELDTTRKVALLOVGTAKVGFVK
                    YSIQSSQTLTLTVTSRAASATLPITVTPVNNKTFPSPVTVYASVIRGASPILR
                    ASVTALIESVNGKTVLELIDNGADATKNDGYSRFTAFDANGRSVIMALGV
                    TSDROARAPKRNKRAYIDGMIEDGEBRNPNRPRTSYQDQOLCFSTSSGSEFVATN
                    VPAARIDLPDPPQITDLKASIQGONLVNLTWAPGDYGRASNYIIMSTIVD
                    LKHENTSLQNTTGLIRKESSEIEFEFEAGNFRNGDTQIFIAIQVDSNLSKEI
                    SNAIVSVFIRPEPIPEDSTPPQDISINSTIPIGILVILKIMKMLGEMOVTLSLH-
                    ORIGIN
BASE COUNT      860 a      718 c      693 g      666 t
Query Match      62.0%; Score 1743; DB 12; Length 2937;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 2161; Conservative 0; Mismatches 595; Indels 21; Gaps 4;
Oy 25 ATGGGCGATTAAAGAGTCTGTGTCATCTTGATCTTCACCTTGTAGAAGGGCCCTG 84
Db 15 ATGGAATCTTTGAGAGTCTGTGTCATCTTGATCTTCACCTTGTAGAAGAGTCTG 74
Oy 85 AGTAATTCATGATTCAGCTGAACAAACAATGGCTATGAAGAGTGTGCTGCAATGCAC 144
Db 75 AGTAGTCCCTCATCCATCACTGAACAAACAAGCGCTATGAGGCGATGTCATCCCATAGAC 134
Oy 145 CCCAATGTGCGAAGATGAACACATCATTCAACAATAAAGACATGGTGACCCAGCA 204
Db 135 CAGAGAGTCCCGGAAGATGAACCCCTCATCAACATTAAGACATGTGTGATCGAGCC 194
Oy 205 TCTCTGATCTGTTGAGCTACAGGAAGCGATTATTTCAAAAATGTGTCATTTG 264
Db 195 TCTCCATCTGTTTAAAGCTACAGGAAAAAGATTATTTCAAAAATGTGTCATTTG 254
Oy 265 ATTCTGAAACATGGAAGCAAAAGGCTGACTATGTGAGCAAAAACCTTGAGACTACAA 324
Db 255 ATTCGGAAGAGTGAAGCAAAAGCCTGAAATAGAGGCCAAACCTTGAAACCTTCAAA 314
Oy 325 AATGCTGATGTTGCTGCTGAGTACTCCTCCAGTAATGATGAGACCTACACTGAG 384
Db 315 AAGGCTGATGTTGCTGCTGATCAACAAACACACCTCTAGGAAAGATGAGCCCTACACGGA 374
Oy 385 CGATGGGCAAGTGTGAGAGAGAGGCTGAAGAGATCCACCTGACTCTGATTTTCATTTGA 444
Db 375 CATATGAGAGCATGTGAGAGAGAGAGATCAGAGATTCACCTGACTCTGATTTTCATTTGA 434
Oy 445 GGAAGAAAGTATGAGTAATGAGACCAAGGTAGGCAATTTGTCCATGAGTGGGCTCAT 504
Db 435 GGAAGAAAGTATGAGTAATGAGACCAAGAGACACTTTGTCCATGAGTGGGCTCAC 494
Oy 505 CTACGATGGGAGTATTTACACAGATACATTAATGATGAGAAATTTCTATTACCAATGGA 564

```



QY	2719	TTTAAATAATGATGGAATGATAGGACAACGCGAGCTGTAATGCGCTAGGCGCAT	2778
Db	2700	CTGAGATTAATGGAATGCGCTAGGCGCAAAATGACGTGATCATTGTTGACCTGAAT	2759
QY	2779	TTTGTGACAGATAAATATAA	2795
Db	2760	TTTCAAGCAAGAAATCA	2776
RESULT	4		
BTU36445			
LOCUS	BTU36445	2984 bp	mRNA
DEFINITION	Bos taurus	calcium-activated chloride channel mRNA, complete cds.	MAM 09-FEB-1996
ACCESSION	U36445		
NID	91184065		
VERSION	U36445.1	GI:1184065	
KEYWORDS			
SOURCE	cow.		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia; Eutheria: Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.		
AUTHORS	1 (bases 1 to 2984)		
TITLE	Cunliff, S.A., Awayda, M.S., Bubljen, J.K., Ismailov, I.I., Arrate, M.P., Berdick, B.K., Benos, D.J., and Fuller, C.M.		
JOURNAL	Cloning of an epithelial chloride channel from bovine trachea		
MEDLINE	J. Biol. Chem. 270 (52), 31016-31026 (1995)		
REFERENCE	96125078		
AUTHORS	2 (bases 1 to 2984)		
TITLE	Benos, D.J.		
JOURNAL	Direct Submission		
DEFINITION	Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics, University of Alabama at Birmingham, BHSB 706, Birmingham, AL 35294, USA		
FEATURES			
SOURCE	Location/Qualifiers		
	1..2984		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/cell_type="epithelial cells"		
	/tissue_type="trachea"		
	19..2730		
	/note="epithelial chloride channel"		
	/codon_start=1		
	/product="calcium-activated chloride channel"		
	/protein_id="AAC48511.1"		
	/db_xref="PID:g1184066"		
	/db_xref="GI:1184066"		
	/translation="MVPRLTYILFLHLHPGKMSVMNINNGYDGVIAINPSYBEDEKIQIKEDVETASTLYLFKATRRYFRVSLIPMTKSKSEYLPQESYDAEVIYANPIKGDDEDTLOYCGKRGQYIHETPNFLITNNLPYSGRAFGVEMALNNGIEFDNGDPPYISIRNMTIATRECSHTGTGNVVKVKGSGSCIRPRDSOGILYEACCTIPEKSPARRESIMFQSLHSTECTCTKTINVAAPNLOKMCAGKSTWYITMNSDTFQNSPTEMANPPTQPTSLKSKQVCLVLDKSGSMSEDRLEPMQAELFLDIQIKSGSLVGVATDSVAEIRNNLTKITDVAENYNTIANPQANGSTJCRGLKAGFAIIIOOSQSTGSEIILITDEDEMIHSCIEVQSGVLIITLAKGSAKELETLSDTGGRHFRFANDINGLTNAFGRISRSQSIQOTIOLESKALATEKRWNGTVPVDSITGDFEFVVTWTKKPEIILDDPKGRKYSIDPKEDKLNHSAIRLIPGIAENG TWYSLINNHASPOLITVYTTTRASTPTTPVYATAMSONTAHESPVIYVAQVSGFLPVGINTVATITETEDGHVTELLMNGAGADVAKNDGYSRTFYDRKNGRSLKVHAERNNNTARLSLRPONKALYIPGYENGLIINPRPEVKDILAAELIEDSRLSGSGSTGAPGPNHPSVLPNKKIIDLAKRFEDHIQISWPAVAVLDRKANSYIIIRISKSLDLOKPDNATLVNTSISLKRKAGSDENFEKPEPIENGITWFIYAVOAINEA NLTSEVNIQAIRKFIEMPEDSVPALGTRKISAINLAIKALMILSTIV"		
BASE COUNT	1027 a	583 c	561 g
ORIGIN		813 t	
Query Match	33.2%	Score 933.4;	DB 3; Length 2984;
Best Local Similarity	61.9%	Pred. No. 1..4e-237;	
Matches 1635; Conservative	0;	Mismatches 966;	Indels 42; Gaps 8;

D	b	7	ACATGCGCAAAAATGAGTCCCTCGCTCGTACGTGATCTCTGTCTCTACTTTGGACACCTCTCG	66
O	y	73	GAAAGGGCCCTGAGTAATTCATCTACCTTCAGCTGACAAACAATGGCTATGAAAGCAATTGTC	132
D	b	67	CTGTG---AATGAAGAAAGTCTCAATGTGTAATTTGATTAACAATGAGATATGATGGCATTTGTC	123
O	y	133	GTTGCATTCGAGCCCATCTGTCAGTGGCCAGAAAGATGAAACCTCATTTCAACAATAAAGGACATG	192
D	b	124	ATTGCAATTAACCTCTAGTGTGCCAGAAAGATGAAAACTCATTTCAAAACATAAAGGAAATG	183
O	y	193	GTGACCCAGGCATCTCTGATCTCTGTTTGAAGCTACAGCAAGAGATTTTATTTTCAAAAAT	252
D	b	184	GTAAGTGAAGCTTCTACTTACTTACTCTGTTCATGACCAACAAGAAAGATTTATTTTCAGAAAT	243
O	y	253	GTTGCCATTTTGAATTCCTGAACATGAGACAAAGCGCTGACTATGTGAGACCAAACTT	312
D	b	244	GTAAGCAATTTTAATTCCAATGACGTGGAGTCAAAATCTGAGTACTTAATGCCAAACAA	303
O	y	313	GAGAGCTTACAAAATATGCTGATGTTTCTGTTGCTGCTGAGTCTACTCTCCAGTAAATGATA	372
D	b	304	GAAATCATATGACAGGAGAAAGTCAATATGTGCTAATCTTACTTAAACATGAGAGTGT	363
O	y	373	CCCTACACTGACGATGGGCAACTGTGGAGAAAGGTGAAAGATCCCACTCTACTCT	432
D	b	364	CCCTATACACTCAATATGGAAGATGTGGAGAAAAGAACAAATATATACATTTTACTOCA	423
O	y	433	GATTCATTCGACGAAAAAGTTAAGTGAATATGACACCAAGATGAGGCATTTGTCCAT	492
D	b	424	AACCTTCTGCTGACTAATAATTTGCCATCTATATGAGGTCCCGAGGACGACATTTGTCCAT	483
O	y	493	GAGTGGGCTCATCTACATGATGGGAGTATTTGACAGTACAAATATATGATAAATTCAC	552
D	b	484	GAGTGGGCCATCTCCGATGGGAAATTTTGAATGATATAGGGGACCAAGCCATTTCTAT	543
O	y	553	TTATCCAAATGG--MAGATACAAAGATGAATGTTTCAGAGATTAATCTGATCAAAAT	609
D	b	544	ATTCGCAAGAAAGAACCTATTGGAAGCAACAAGATGTTCAACTCATATTTACTGCTACTAAT	603
O	y	610	GTAATTAAGAAGTCTCAGGGAGGCACTGTTACACAAAAGATGCATTTCAATTAAGTA	669
D	b	604	GTGATGTGCAAAATCCGAGGAGGACAGCTGTATTAACAAGGCATTCGAGAGTACTCACAG	663
O	y	670	ACAGAGCTATATGAAAAAGAGATGAGTTGTTCTCCAAATCCCGCAGAGGAGAGAAGCT	729
D	b	664	ACAGGCTGTATGAGCAAAATGTACATTTATCTCCAGAAAAATCCAGACTGCCAAGGAG	723
O	y	730	TCTATTAATGTTTGCAACAATGTTGATTTCTATAGTTGAATTTCTGTACAGAACAAAAACAC	789
D	b	724	TCCATCATGTTTATGCAAAATCTCTCATCTCTGTGACTGAATTTTGTACAAAAAACACAC	783
O	y	790	AACAAAGAGCTCCAAACAGCAAAATCAAAAATGCAATCTCCGAAGCACTATGGGAATG	849
D	b	784	AATGTGGAAGCTCCAAACCTCAAAACAAATATGGAATGGCAAAAAGCAGATGGAGATGA	843
O	y	850	ATCCGATATCTGAGACTTTAAGAAACCACTCTATGACA-----ACACAGCCACCA	903
D	b	844	ATCATGAACCTTACTGACTTTTCAAAATACATCTCCCATGACAGAAATGAATCCACCACT	903
O	y	904	AATCCACCTTCTCATTTGCTGACAGATTGGACAAAAGATTGTGTAGTCTTGACAAA	963
D	b	904	CAACCTCATATTTTATTTGCTCAAGTCCAAACACACAGAGTATGTTGTGGTACTTATATA	963
O	y	964	TCTGGAAGCATGGGAGATGGTAAACCGCTCAATGCACTAATCAAGCAGAGCAGCTTTTC	1022
D	b	964	TCTGGAAGCATGTTTAGAAGATGCTCTCTTTCGATGAATATCAAGCACAGAAATTTTC	1023
O	y	1024	CTGTGCGACAGATTGAGCTGGGGCTCTGTGGTGGAGATGATACATTGACAGTCTGCC	1083
D	b	1024	TTGATTCAAATTAATTAAGAAAGGATCTCTGTGGATGGTATACATTTTACAGTGTTCCT	1083
O	y	1084	CATGTCAAAAGTGAATCTATACAGATTAACAAGTGGCAGTGCACAGGACACACTCGCCAAA	1143

Db 1084 GAATCCGAATTAATCTAACAATAAAGTATGATATGTTATGAAATATCACTGCA 1143  
 Qy 1144 AGATTACCTGACGACGCTTCAGGAGGAGCGCTCCATGCGACGGCTTCGATCGGCAATT 1203  
 Db 1144 AATCGCTCAGAGAGCTAATGAGGAGCTTCAATTTGTAGGGGCTCAAGAGCAATTT 1203  
 Qy 1204 ACTGTATTAGGAGA---AATATCCAACTGATGAGTCAATTTGCTGCTACAGAT 1260  
 Db 1204 CAGGCAATTTATTCAGAGTCAACAGAGCACTTCTGCTTGAATCAATCTATACCGAT 1263  
 Qy 1261 GGGGAGACACACTATTAAGTGGGTCTTAACGAGCTCAACAAAGGCTGATCATC 1320  
 Db 1264 GGGGAGAGTAATGAATATCACTATGATGAGGAGTAAACAAAGGCTGATCATC 1323  
 Qy 1321 CACACAGCTGCTTGGGGGCTCTGACAGTCAAGAACTAGAGAGCTGTCCAAATGACA 1380  
 Db 1324 CACACATTTGCCCTGGAGCTTCTGTGCAAGAACTGAGACACTGTCAAGATATGACA 1383  
 Qy 1381 GAGGTTTACAGACATATGCTTCAATCAAGTTCAGAACATGGCTCATTTGATGCTTT 1440  
 Db 1384 GAGGACATCGGTTTATGCAATTAAGACATA---AATGGCTTACTATATGCTTC 1437  
 Qy 1441 GGGGCTTTCATCAGAAATGAGGCTGTCTCAAGCTCCATCCAGCTTGAGATAG 1500  
 Db 1438 AGTAGAATTTCACTAGAACTGAGTATCAGTACAGCACTATTCAGTTGGAAGCAAA 1497  
 Qy 1501 GATTAACCTCCAGAACACGCGAGTGAATGAGCAAGTGAATGAGACACCGTG 1560  
 Db 1498 GCTTGGCAATACAGAAAGAAATGGTAAATGGTACGGTCTGTGATGATCAAT 1557  
 Qy 1561 GGAAGAGCACTTGTCTTATCAGTGGACAAAGCAAGCTCCCAATCTCTCTGG 1620  
 Db 1558 GGAATGACCTTTCTTGTGTGACATGACATTAAGGACAGAAATTTCTCCAG 1617  
 Qy 1621 GATCCAGTGCAGAGCA---AGTGGCTTTTATGAGCAAAACACCAAA 1671  
 Db 1618 GATCCAAAGAAAGAAATTAACCTCAGATTTCAAAAGAAATAGCTAAATATTCAT 1677  
 Qy 1672 ATGGCTTACCTCCAAATCCAGGCTTGTAGGTTGGCACTTGGAATACAGTCT--- 1728  
 Db 1678 TCTCTGCTTCTGCAATACCTGATTTGACAGAGCAGTGTGATGATGATGATCTTA 1737  
 Qy 1728 ----GCAAGCAAGCTCAAAACCTGACCTGCTCAGCTCAGCTCCGCTGCTCAATGCT 1782  
 Db 1738 AATATATACGACGCTCAATTAAGTACAGTGCAGTGCAGTGCAGAGAGAGCT 1797  
 Qy 1783 ACCCTGCTCCAAATTAAGTACTTCCAAACAGCAAGACACCAATTCGCCAGC 1842  
 Db 1798 ACTACACCCCAAGTACGACAGCTCAGTACAGTCAAAATTAAGACACTACCTAGC 1857  
 Qy 1843 CCTGTGAGTTTATGCAAAATATTCGCCAGAGAGCTCCCAATTCAGAGGCAAGTGC 1902  
 Db 1858 CCAGGATTTGTTATGACCAAGTCAAGGTTTTGCTGCTGAGGAAATTAATGA 1917  
 Qy 1903 ACAGCCCTGATTTGAATCAAGTGAATGGAAGAAACAGTTACTTGGAATCTGATATGA 1962  
 Db 1918 ACTGCAATTAAGAACTGAAGATGCGCACTCAAGTACACTGAGCTGTGGACAAAGT 1977  
 Qy 1963 GGAGGCTGATGCTACTAAGATAGCTGTCTCACTCAAGGATTTTCAACTTAATGAC 2022  
 Db 1978 GCGAGTGTGATCTCTCAAGATATGCAATGCTATTCAGATATTTTACGATACCGT 2037  
 Qy 2023 ACAGATGTAGATACAGTATAAAGTGGGCTCTGGGAGGAGTAAAGCAGCGAGCAGG 2082  
 Db 2038 GGAATGTAGATAGTAAAGTAAAGTACATGCCGAGGAGAAACACACGCTAGGCTA 2097  
 Qy 2083 AGAGTATACCCAGAGAGTGGAGCTGTATCATCTGCTGATGATGAGATGATAA 2142  
 Db 2098 AGTTAAGACAAACACAGAACAAAGCCCTGTATATACAGGCTAATATGAAATGTAAA 2157  
 Qy 2143 ATACAATGAAATCCACAGAACTGAATTAATAGGATGATGTTCAACACAGCAAGTG 2202  
 Db 2158 ATATATCTGAACCCACAGCACTGAAGTCAAGATGACCTGCAAAAGCTGAATATGAA 2217

Qy 2203 TGTTCACAGACATCCTCGGAGGCTCATTTGTGCTTGTATGATCCAAATGCTCC 2262  
 Db 2218 GACTTTAGACAGACATCCTCGAGGGGCTATTTACTATACAGAGCTCTCCTGTATAT 2277  
 Qy 2263 ATACCTGATCTCTCCACCTCGGCCAAATCAGCACTGAAAGCGGAATTCACGGGGC 2322  
 Db 2278 CACCTTCTGTGCTCCACCAATTAATTAAGATTTAGACCTTGAGCAAAATTTAAAGAA 2336  
 Qy 2323 ACTGCAATTAATCTGACTTGGACAGCTCTCGGGATGATTAATGACATGGAACAGCTCAC 2382  
 Db 2336 -ATCACATTTCACTTCAATGAGACAGCCCTCGATGATTAAGGAAAGCCAC 2394  
 Qy 2383 AAGTATATCTGAAATTAAGTAAAGTATTTCTGATCTCAAGACAAAGTCAATGAATCT 2442  
 Db 2395 AGCTACATTAATTAAGATTAAGTAAAGTATTTCTGATCTCAAAAGATTTGACAAATGCT 2454  
 Qy 2443 CTTCAGTAATTAATCTGCTCATCCCAAGAGCAAGCACTGAGAGAGCTTTTGG 2502  
 Db 2455 ACTTATGATTAATCTTCAATGAGACAGCCCTCGATGATTAAGGAAATTTTGA 2514  
 Qy 2503 TTTAAACAGAAACATTAATCTTTGAAATGCGACAGATCTTTTATGATGATTCAGGCT 2562  
 Db 2515 TTTAAACAGAAACCTTTTGAATTAAGAAATGCGACCACTTATATGACAGTCCAGCC 2574  
 Qy 2563 GTTGAATAGTGTGATGATGAAATCAATCAATGACAGATCTTTGTTAT 2622  
 Db 2575 ATCAATGAAGCAATCTCACCTCAGAGTTTCTTACATGCGCAAGCAATCAAGTTAT 2634  
 Qy 2623 CCT 2625  
 Db 2635 CCT 2637

RESULT 5  
 AF001261 3317 bp mRNA MAM 19-NOV-1997  
 LOCUS Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF001261  
 NID 92623762  
 VERSION AF001261.1 GI:2623762  
 KEYWORDS  
 SOURCE  
 ORGANISM Bos taurus.  
 Bos taurus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.  
 REFERENCE 1 (bases 1 to 3317)  
 Elble, R.C., Widom, J., Gruber, A.D., Abdel-chany, M., Levine, R.,  
 Goodwin, A., and Pauli, B.U.  
 Cloning and characterization of Lu-ECAM-1 suggest it is an  
 endothelial chloride channel  
 Unpublished  
 JOURNAL 2 (bases 1 to 3317)  
 REFERENCE Elble, R.C., Widom, J., Gruber, A.D., Abdel-chany, M., Levine, R.,  
 Goodwin, A., and Pauli, B.U.  
 Direct Submission  
 Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
 Ithaca, NY 14853, USA  
 FEATURES  
 source  
 1..3317  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /cell\_type="aortic endothelial cells"  
 /clone="1"  
 /tissue\_type="lung"  
 /note="endothelial adhesion molecule; chloride channel"  
 /codon\_start=1  
 /product="Lu-ECAM-1"  
 /protein\_id="AA86529.1"  
 /db\_xref="PID:92623763"

CDS



Db 1834 CAGAGATCTCTATCTATACCCCGATTAATTCGACAGCTCAGCTGATGCTACATTAACGAC 1893  
Qy 1832 AATTCGCCAGCCCTCTGGTGTATTTATGCAATATTCGCCAGAGACCTCCCAATTCCTCA 1891  
Db 1894 ATTATCTAGAGCCCAATGATGTTTATGCAAGTCAAGTCAAGGTTTTCCTGATCTGG 1953  
Qy 1892 GGGCAGGTCTCAGAGCCCTGATGTAATCAAGTAATGGAAGAAAGTTCCTTGAACATC 1951  
Db 1954 GAATCAGGTATTAATGATCAATTAAGAAACCAAGGATGACATCAATGATGAGCTCT 2013  
Qy 1952 TGGATTAATGAGCAGGTGCTGATGCTACTAAGATGAGGCTGCTACTCAAGTATTTCA 2011  
Db 2014 GGGACAAATGCTGAGGTCTGATGCTAGTCAAGAAATGATGCTACTCAAGATCTTTA 2073  
Qy 2012 CAACTTAATGACAGATGCTGATGCTAAGTAAAGTGGGCTCTGGAGAGGATTAAG 2071  
Db 2074 CAGATTAATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2133  
Qy 2072 CAGCAGAGGAGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2131  
Db 2134 CGGCTAGGCTAAATTAATTAAGAACCAAGAGTCTATATGCTCCAGGCTGATGATG 2193  
Qy 2132 AGAATGATGAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2191  
Db 2194 AAAAGGATGAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2253  
Qy 2192 ACAAGCAAGTGTGTTTACAGACAGATCTCGGAGGCTCATTTGTGCTGTGATGCTC 2251  
Db 2254 CTAATATGAGAACTTTGACAGACTACCTCTGAGAGGCTCATTTGATGATGAGAGCTC 2313  
Qy 2252 CAAATGCTCCCA---TACTGATCTCTCCACCTGAGGCTCAACGACCTGAGAGGCG 2308  
Db 2314 CTTCTCTGCTGATATCACTCTGCTGCTCCACCAAGTAAATTAAGATCTTTGAGGCTA 2373  
Qy 2309 AAATTCACGAGGAGGAGTCTCATTAATCTGATGATGATGATGATGATGATGATGATG 2368  
Db 2374 AGTTCAAGAAAG---ATTATATTAATCACTTCAATGAGACAGCCCTGACATGCTGATA 2430  
Qy 2369 ATGGAACAGCTCAGACAGATTAATCAATTCGAAATAGTAAAGTAAATCTTCTGATG 2428  
Db 2431 AAGGAAAGGCAAGAGCTCATTAATTAAGAAATAGTAAAGTAAAGTAAAGTAAAG 2490  
Qy 2429 AGTTCAATGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2488  
Db 2491 ATTTGACAAATGCACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 2550  
Qy 2489 AGGAGTCTTTTGTGTTTAAACCAAGAAACATTAATTTGAAATGAGCAGATCTTTTCA 2548  
Db 2551 AAGAAATATTTTAAATTTAAGCAGAAACATTTAGATGATGAAATGAGCAGAAATTTCTATA 2610  
Qy 2549 TTGCTATTCAGGCTGTGTAAGTGAATGATGATGATGATGATGATGATGATGATGATG 2608  
Db 2611 TTTCAGTCCAAAGCCTCAACGAAAGCAATCTCATCTCAAGAGTCTTCTCATATTGTA 2670  
Qy 2609 TATCTTTGTTTATCTCTCAG 2629  
Db 2671 CAATCAAAATTAATCTCTTAC 2691

RESULT 6  
AF047838 3022 bp mRNA ROD 26-NOV-1998  
LOCUS AF047838 3022 bp mRNA ROD 26-NOV-1998  
DEFINITION Mus musculus calcium-sensitive chloride conductance protein-1  
(mclcl1) mRNA, complete cds.  
ACCESSION AF047838  
NID 93925280  
VERSION AF047838.1 GI:3925280  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3022)  
AUTHORS Gandhi, R., Eble, R.C., Gruber, A.D., Schreier, K.D., Ji, H.L.,  
Fuller, C.M., and Pauli, B.U.  
TITLE Molecular and functional characterization of a calcium-sensitive  
chloride channel from mouse lung  
JOURNAL J. Biol. Chem. 273 (48), 32096-32101 (1998)  
MEDLINE 99041980  
REFERENCE 2 (bases 1 to 3022)  
AUTHORS Gandhi, R., Eble, R.C., Gruber, A.D., and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (12-FEB-1998) Vet. Pathology, Cornell Univ., VMC C4-146,  
Ithaca, NY 14856, USA  
FEATURES  
source location/Qualifiers  
1..3022  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/tissue\_type="whole lung"  
1..3022  
/gene="mclcl1"  
18..2726  
/note="calcium-activated chloride channel; Alternative  
chloride channel to CFTR that may influence severity of  
Cystic Fibrosis in different tissues; mclcl-1"  
/product="calcium-sensitive chloride conductance  
protein-1"  
/protein\_id="AAC79982.1"  
/db\_xref="PIR:G3925281"  
/db\_xref="GI:3925281"  
/translation="MVPGLQVLFLLHLHONTSSVHVLNSGVEGVIAIRPSVE  
DEHLIPSKIMVVOASLYFEASGRYFENSLIVPMTKSPETVLMRESYDAD  
VYADPRLQGDPPYLTQYQCGDROGYHFTFNFLVHNEERGVTRACRSTVRL  
RWGFFEDINVDPPYMSKNTIATRTSTITGVNVLNDEKGVTRACRSTVRL  
YERKCTIPDKIQTAGASIMFQNLNSVFECEKHNNAAPNQNMCNRSITVDI  
KTSADFQNAAPMGTEAPPEPTSLKSRVVCVLYDKSGMDKEDRLIRMQAEL  
YLQIVYKESWGLVTFDSAHIONYLKITSSDOKITANLPQAGSGSTICHLQ  
AGFOATISDSOSTSGSEIVLLTGDENDGIRSCFEVRSGLIHTLALGSAARELT  
LSDMTGLRFRANKDUNSLDARSISTSGSYSOALQLESAPFVRAQATNGIYP  
LDSTVGNDITFVITWYKPEIILQDKKKTITSDQDQKLRSLALIPGTAETG  
TWYSITGTSOLITWTVTRAPSPTEPLATAHMSQTAQYPSRMIVARVSQGL  
PYGANTALIEAHEHGVLELMDNAGATVKNQIYRTFEDYDGNRYSILKAV  
QAQRNKTRLILROKNSLYIPGVENKILVNPDPVQBEALTEADENRYTSGS  
FTYSAGPDQDRAVPSPKVTDLAEFTIGDYHILMTARGLVLRARHRYIIRMSQ  
HPILDQEDFNATLVNASSLIPKESKEFKRPERFKANGIQLYIALQADNEAL  
TSEVSNIAQVAKLTSLSDSLALGDDISALSMITWGLTVLFSILN"

BASE COUNT 923 a 710 c 661 g 728 t  
ORIGIN

Query Match 30.7%; Score 863; DB 12; Length 3022;  
Best Local Similarity 61.18; Pred. No. 7.2e-219;  
Matches 1568; Conservative 0; Mismatches 960; Indels 39; Gaps 9;

Qy 91 TCACATTCAGCTGACGACCAATGCTATGAGCAATGCTGTCATGACCCCAAT 150  
Db 81 TCATAGTGTGATCTCAGACGATGATGATGATGATGATGATGATGATGATGATGATG 140  
Qy 151 GTGCCAGAAAGTGAACACATGATTAACAAATTAAGAGATGATGATGATGATGATGATG 210  
Db 141 GTGCCGAGAGAGCAAGAGCTATCCCAAGCAATTAAGAGATGATGATGATGATGATGATG 200  
Qy 211 TATCTGTTGAAGTACAGGAAAGCAATTTATTTCAAAATGTTGCCATTTGATTCCT 270  
Db 201 TACCTGTTGAAGCAGCAAGCAAGATTTATTTCAAAATTAAGATGATGATGATGATGATG 260  
Qy 271 GAACATGAGAGCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 330  
Db 261 ATGACCTGGAAGTCAAAACCTGATGATGATGATGATGATGATGATGATGATGATGATG 320  
Qy 331 GATGTTCTGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 390  
Db 321 GACGTCATAGTGTGGAGATCTCAGCTGACATGAGAGCAGCCCTACACCTTCAGAT 380





Db 2529 ATAGCAATGGCATCTACCTACTTTCATTCAGGACAGCAATGAAGCCAGTCTCAC 2588  
QY 2584 TCAGAAATATCCAACTGACAGAGTATCTTGTATTATCTCCACCA 2630  
Db 2589 TCTGAGGTCTCCAACTGACAGAGTCTGCAAGCTTACTCTCTGA 2635

RESULT 7  
AF052746 3471 bp mRNA ROD 08-Apr-1999  
LOCUS Mus musculus chloride channel cacc mRNA, complete cds.  
DEFINITION AF052746  
ACCESSION AF052746  
NID 93560546  
VERSION AF052746.1 GI:3560546  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3471)  
AUTHORS Romio, L., Musante, L., Cinfl, R., Serl, M., Moran, O., Zegarra-Moran, O.  
and Galietta, L.J.V.  
TITLE Characterization of a murine gene homologous to the bovine cacc  
chloride channel  
JOURNAL Gene 228 (1-2), 181-188 (1999)  
MEDLINE 99173882  
REFERENCE 2 (bases 1 to 3471)  
AUTHORS Romio, L., Musante, L., Cinfl, R., Moran, O., Serl, M. and  
Galietta, L.J.V.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto  
Glennina Gaslini, Genova I-16148, Italy  
LOCATION/Qualifiers  
SOURCE 1. 3471  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/db\_xref="dbEST:AA245677"  
47..2752  
/codon\_start=1  
/product="chloride channel cacc"  
/protein\_id="AAC35003.1"  
/db\_xref="PID:93560547"  
/db\_xref="GI:3560547"

CDS  
/translation= WPGDLYLFLTHLONTSSVHLNSNCEYVAINPSVE  
DERLIPSIKEMVTOASTYLEASQRYFNISLLVPMYKSKREYIMPRESDKAD  
VIVADPHLOHDDPYTLQYGGCGDQGYIHTPNFLTDNLRIVGPRGYFVEMAH  
RWGVEEDVNDOPPYMSRKNTEATRCSTRTNVVHNCERGCVRACRDRSKTL  
YEPKCTFIPKIOFAGASIMEMONLSEVECTEKNNAEAPNLOKNCRRSDWVY  
KTSADPQNPMPRGTEAPPEPTESLSRVVCVLYDKSGSMKEDRLRMNQAEL  
YLIOIVKESWGLVTFDSAHIONTYIKITSSDYKITANLPQASGCTSTGCIQ  
AGFOATISSDOSTSSEIILTDGEDNGINSCEFAVSQAIIHTIALGSAAELET  
LSDMTGGLREYANDLSLIDAFSRISSTSGVSQAOLESKAFDVRAGAMINGTP  
LDSVNDGTFFVITWYKKEPIILDPEKGYSDPDODKLNRSARLIQPAETG  
TMTYSITGTSOLITMTVTTRASPTMEPLATTHMSOAYOVSMTIVARYSGFL  
PVGANNTALIEAHGHQVTELEMDNAGDVTYNDIYRERYFDYGNRSYKAYV  
QAOENKTRLSLRKNSLYIPGVENKRYVLPNRPDVOEALIEAIEDENRYTSQGS  
FTVSGAPDGDHARVPPSKVTDLEAFIDGYHLTWTAQKVLNDRARHRYIIRMQ  
HPTLDQDFNNATLVNASLSLIPKENGSKETFEKFEFKIANGIOLIAIOADNEAST  
TSEVSNIAQVKTSLSDSISALGDSIAIKMTIGLIVIFENSLN"

BASE COUNT 1096 a 777 c 736 g 862 t  
ORIGIN

Query Match 30.7%; Score 863; DB 12; Length 3471;  
Best Local Similarity 61.1%; Pred. No. 7.2e-219;  
Matches 1568; Conservative 0; Mismatches 960; Indels 39; Gaps 9;

QY 91 TCACATCATGCTGACACAACTGCTATGAAGCATTTGCTTGCATGACCCCAAT 150  
Db 110 TCATATGCTGCTCAACAGCAATGATACGAGGCTGTGCATTTGACATTAACCCAGT 169  
QY 151 GTGCCAGAAAGTGAACACTCATTCACAAATTAAGGACATGGTGAACCCAGCATCTGTG 210

Db 170 GTCCCGGAGGACGAAAGGCTCTATCCAGCATTAAGGAATGTAACCTCTTACC 229  
QY 211 TATCTGTTGAAGCTACAGGAAAGCATTTTATTTCAAAAATGTTGCATTTTATCT 270  
Db 230 TACCTGTTGAAGCAGCAGCAAGGAAGTTTATTTCAAAAATGTAACATTTATGTCGG 289  
QY 271 GAAACATGGAAGCAAAAGGCTGACTATGTGAGACCAAACTTGAGACTTACAAAATGCT 330  
Db 290 ATGACCTGGAAGTCAAAACCTGACTTATTAAGCAAAACAGAGATGTAAGCAAAAGCA 349  
QY 331 GATGTTGCTGTTGCTGAGTCTACTCTCCAGGTATGATGAAGCACTACAGTACAGATG 390  
Db 350 GACGTGATAGTGGCGGATCTCTACCTGCAACATGAGACGACCCCTTACACCTTCAGTAT 409  
QY 391 GGCAACTGTGAGAGAAAGGCTGAAAGGATCCACCTCACTCTGATTTTCATTTGACAGAAA 450  
Db 410 GGACAGTGTGGGAGACAGAGGACATACATACACTTCACTCCAACTTCTACTGATGAT 469  
QY 451 AAGTTAGCTAATTTGACACCAAGAGTGGCATTTGTCATGATGAGTGGCTCATCTACA 510  
Db 470 AACTTGGGTATCTACGAGACCCCGAGGACAGATCTTGTCCATGAGTGGCCCATCTCCG 529  
QY 511 TGGGAGATTTGACGAGTACATTAATGATGAGAAATTTCTACTATC---CAATGGAAGA 567  
Db 530 TGGGAGATTTGATGAGTATTAACGTGACCCAGCTTTCTACATGCTTAGAAGAACACT 589  
QY 568 ATACAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 627  
Db 590 ATAGAGCAACAAAGTGTCTCCACAGATGATGATGATGATGATGATGATGATGATGATGAT 649  
QY 628 GGAGGAGCTGTTCACCAAAAGATGACATCTCATATAAGTAAGAGACTTATGAAAAA 687  
Db 650 AGAGGCACTGTGTCACAAAGGCGTGGCGGTGCTGACGAGGACGCGGTATGAAAGCC 709  
QY 688 GGATGTGAGTGTGTTCTTCATCCCGCAGAGAGGAGAGGCTTTATATGTTTGCACAA 747  
Db 710 AAATGTACATTTATCCAGCAAAATATACAGGCGCTGGGCGCTCCATTAATGTTATGCA 769  
QY 748 CATGTGATCTATAGTGAATTTCTGTACAGAACAAACCAACAAAGAGCTTCAAC 807  
Db 770 AACCTTATTTCTGTGTTGATTTTGCACAGAAAAAACCAATGACAGAGCCCAAC 829  
QY 808 AAGCAAAATCAAAATGCAATCTCCAGACACATGGAAGATGATGATGATGATGATGATGAT 867  
Db 830 CTACAAAAAAGTCAACGCAAGCAAGGATGTAACAGAGCTGCTGAC 889  
QY 868 TTAAAGAAAACCACTCTATG-----ACAACAGCCACCAATCCACCTTCTCATG 921  
Db 890 TTTCAGAAATGCCCCCTCCCATGAGAGAACAGAACCCCTCTCCACCTATTTTCACTG 949  
QY 922 CTGCAATTTGACAAAGATTTGTGTTAGTCTTACAAATCTGGAAGCATGCGACT 981  
Db 950 CTCAAGTCCAGAAAGGACAGTGTGCTGCTGCTGATTAATCTGGAAGCATGACAAA 1009  
QY 982 GGTAAAGCCCTCATGACATGATCAAGCAAGGACGCTTTTCCGTGCGACAGAGTGAAG 1041  
Db 1010 GAAAGCGCTCTTATTAAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1069  
QY 1042 CTGGGAGCTGAGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1101  
Db 1070 AAGAGTCTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1129  
QY 1102 ATACAGATTAACAGTGCAGTACAGAGGACACACTGCGCAAAAGATTAACCTGACAGACT 1161  
Db 1130 ATAAATTAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1189  
QY 1162 TCAGAGGAGAGTGCATCTGACAGGCGCTTGCATGCGCATTTTACTGATTA---GGAAG 1218  
Db 1190 TCTGTGGAACCTTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1249  
QY 1219 AAATATCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1278



Db 1250 GACCAGACACTTCCGCTTCTGATGCTATTCCTGACAGATGGGGAAGATAATGATA 1309  
 QY 1279 AGTGGGTGCTTTAAGCAGATCAAAAGTGGTCCATCATCCACACAGTCCGTTGGGG 1338  
 Db 1310 CGTTCCTGCTTGAAGGCGCTCTCGAGCGGTGCTCATTCACACATCCGCTCTGGGG 1369  
 QY 1339 CCCTTCGACACTCAAGACACTAGAGAGCTCCAAAATGACAGAGGTTTACAGACTAT 1398  
 Db 1370 CCTCGGCTCCGAGAACTGAGACTCTGCGACATGACAGAGGCTTCGTTCTAT 1429  
 QY 1399 GCTTCAGATCAAGTTCGAAACATGGCCCTATGATGCTTTGGGGCCCTTTCATCAGA 1458  
 Db 1430 GCCAACAAAGACCT-----AAACGCTTATGATGCTTCATGATGATTTATCTACA 1483  
 QY 1459 AATGAGCTGTCTCTGACGCGCTCCATCCAGCTTGAGATGAGGAGATTACCTCCAGAAC 1518  
 Db 1484 AGTGGCAGGCTCTCCAGAGGCTCTGACAGTTCGAGAGCAAAAGCCTTCGATGTCAGACA 1543  
 QY 1519 AGCCATGATGATGACAGACAGTATGTCGACAGACCGCTGGGAAAGACACTTTGTTT 1578  
 Db 1544 GGGGATGATGATTAACGCTACAGTACCTCTGGACAGTACCTCGGCAACGACGCTCTTT 1603  
 QY 1579 CTATACCTGAGCAACGACGCTCCCAATCTCTGAGGATCCAGTGGACAGAAAG 1638  
 Db 1604 GTTATCAGCTGATGATGATAAAAGCCAGAAATATTTCTTAAGATCCAAAGGAAAAA 1663  
 QY 1639 CA-----AGTGGCTTTTGTAGTGGACAAAACCAAAATGGCTTACCTCCAAATC 1689  
 Db 1664 TATACAACTCAGATTTCCAAAGATGATTAACCTAAACCTCGGCTGCTGATCACTCAATA 1723  
 QY 1690 CCAGGATGCTAAGTTGGCACTTGGCAATTAAGTCT---GCAAGCAGCTCACAACC 1746  
 Db 1724 CCGGGGCTGACAGACAGAGTACTTGGACTTACAGATCAAGGATCAAGGCTCAGTTG 1783  
 QY 1747 TTGACCTGACTGTCACTGCTCCGCTGCTCAATGCTACCTGCTCCAAATACAGTACT 1806  
 Db 1784 ATTACATATGACATGACACTGACACTGACAGAAAGTCCACCACTGAGACACTCTGCTACT 1843  
 QY 1807 TCCAAACGACAGACAGACAGCAAAATTCGCCAGCCCTCTGTGATGTTATGCAAAATAT 1866  
 Db 1844 GCTACATGATGATGACAGACAGCCAGTACCTTACCCGAGTATGATGATGACAGGCTC 1903  
 QY 1867 CGCAGAGAGCCCTCCCAATCTCAGGCGCAGTGTACAGACCTGATGATGATGATGAT 1926  
 Db 1904 AGCAGAGATTTTGGCTGCTGCTGAGGACCAAGTGTACACACCTCTCATAGAGGTGAACAT 1963  
 QY 1927 GGAAGAAAGTTCCTGGAATCTGGAATCTGATATGAGCAGAGTGTCTGATTAAGAT 1986  
 Db 1964 GACATCAAGTCACTTGGAGCTGTGGAGCAATGGGCGAGTGTCTGATCTTTAAAT 2023  
 QY 1987 GAGGCTGTCTACTCAAGATATTCACAACTTATGACAGCAAGATGATGATGATGAT 2046  
 Db 2024 GATGGATCTACACAGATATCTTACAGATATGATGAAATAGTATGATGATGATGAT 2083  
 QY 2047 GTGCGGCTCTGGAGAGATTAACGACAGCAGAGAGTGTATACCCAGCAGAGTGA 2106  
 Db 2084 GTGCGGTGCTGAGCACAAGAAACCAAAACGAGCTGAGCTTAAGA---CAGAAGAACAG 2140  
 QY 2107 GCATCTATACATCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2166  
 Db 2141 TCTTTATATATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2200  
 QY 2167 GAAATTAATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2226  
 Db 2201 GATGCTCAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2260  
 QY 2227 GGTCTATT---TGTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 2283  
 Db 2261 GGTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2320  
 QY 2284 GGCATTAATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2343  
 Db 2321 AGTAAAGTCAAGACCTGAGGCTGAGTATGATGATGATGATGATGATGATGATGATGAT 2377

QY 2344 ACAGCTCTGGGGATATATATGACCATGGAACAGCTCCACAGATATATTCATTCGAATAGT 2403  
 Db 2378 ACAGCTCTGGGGATATATATGACCATGGAACAGCTCCACAGATATATTCATTCGAATAGT 2437  
 QY 2404 ACAAGATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2463  
 Db 2438 CAGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2497  
 QY 2464 CTCATCCAAAGAGAGCAACCTCTGAGAGTCTTTTGTGTTTAAACAGAAACATTACT 2523  
 Db 2498 CTGATACCTTAAGAGAGCTGCTCAAAAGATTTTAATTCAAACAGAAACATTACT 2557  
 QY 2524 TTTGAAATGACAGATCTTTTCAATGCTATTCAGGCTGTGATGATGATGATGATGAT 2583  
 Db 2558 ATAGCAAAATGAGCATCCAGCTTATTCGATGATGATGATGATGATGATGATGATGAT 2617  
 QY 2584 TCAGAAATATCCAACTTCAGAGATATCTTTGTTATTCCTCCACA 2630  
 Db 2618 TCTGAGGCTCCAACTTCAGAGATGATGATGATGATGATGATGATGATGATGATGAT 2664

RESULT 8  
 AF043976 3415 bp mRNA PRI 07-APR-1999  
 LOCUS AF043976 Homo sapiens ClCA homolog (hCLCA3) mRNA, complete cds.  
 DEFINITION AF043976  
 ACCESSION AF043976  
 NID 94572288  
 VERSION AF043976.1 GI:4572288  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 3415)  
 TITLE Gruber A.D. and Pauli, B.U.  
 JOURNAL Molecular cloning and biochemical characterization of a truncated,  
 secreted member of the human family of Ca2+-activated Cl- channels  
 Biochim. Biophys. Acta 1444 (3), 418-423 (1999)  
 REFERENCE 99196715  
 MEDLINE 2 (bases 1 to 3415)  
 JOURNAL Gruber A.D., Eble, R.C. and Pauli, B.U.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of  
 Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

FEATURES  
 source 1..3415  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..3415  
 /gene="hCLCA3"  
 19..807  
 /codon\_start=1  
 /product="ClCA homolog"  
 /protein\_id="AA023734.1"  
 /db\_xref="GI:4572288"  
 /db\_xref="PIR:G4572289"  
 /translation="MVFSLKVLFLSLILSPYLSKSLVTLNNNGYDGLVIAINPSVE  
 DEKLONIKEMTTEASTHLFHTKRAAFRANVSLIIPMTYKSKSYLLPKOTYQAD  
 VIVADLYLKIDPTLIDYGGGDKGQIHFHPNLLTNLNLATIGPRGVFVGHVHL  
 RMGVDEYNVDPEYISRNNTETASTRTIVYVNLCKGASCIARFRRDSOTGL  
 YEAKCTFIPKRSOTAKESIVMONTDSYTECTEKTENKRAPNL"

BASE COUNT 1149 a 625 c 630 g 1011 t  
 ORIGIN  
 Query Match 29.5%; Score 831; DB 42; Length 3415;  
 Best Local Similarity 59.9%; Pred. No. 2.3e-210;  
 Matches 1602; Conservative 0; Mismatches 980; Indels 93; Gaps 9;

QY 46 GGTGTTATCTTATCTTACCTTCTAGAGAGGCGCTAGATATTCATCTACGCTG 105  
 111 11111 11111 11111 11111 11111 11111 11111 11111 11111

Db 37 GTGATTCCTCTCTATCCCTGCTCTCTGCGCTGTATGAAAAGCTCACTGTAATTG 96  
Qy 106 AACACAATGGCTATGAAGCATGTGCTGTGCAATCGACCCCAATGTGCGACAGAATGAA 165  
Db 97 AATAACATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 156  
Qy 166 ACACATTCATCAACATAAAGACATGTGACCCAGCGCATCTCTGTATGTTGTAAGCT 225  
Db 157 AACATCATTCACAAACATAAAGAAATGTACTGAAAGCATCTACTACCTGTTTCAATGCG 216  
Qy 226 ACAGAAAGCGATTTTATTTTCAAAAATGTGCAATTTTGTCTGAAATCATGAAAGACA 285  
Db 217 ACCAACAAGAGCTTATTTTCAGATGTAAAGCATTTTATTTCCAAATGACACAAATGA 276  
Qy 286 AAGGTGACTATGTAGACAAACTGTGACCTACAAAATGCTGATGTTGCTGCT 345  
Db 277 AATCTGAGTACTTATCCCAACAAACAAATATGACCAAGGACAGATGTCTAGTTGCT 336  
Qy 346 GAGTCTACTCCCTCCAGTATGATGAACCTCTACCTGACGACATGGGCACTGTGAGAG 405  
Db 337 GATCTTACCTGAATACGAGATGATCCCTATACACTTCAATATGAGACATGTGGAGAT 396  
Qy 406 AAGGTGAAGAGATCCACCTCACTCTGATTTTCATTTGACGAAAAAGTTAGCTGAAT 465  
Db 397 AAGGACATATATATACATTTTACTCCAACTCTGTTGACTAATAACTTGCTACCTAT 456  
Qy 466 GAGCCACAAGTAGGCGCATTTGTCATGATGAGGCTCATACGATGGGAGCTATTTGAC 525  
Db 457 GGGCTTCAGAGTAAAGTATTTGTCATGGGTGGGCCATCTCCGGGGAGTATTTGAT 516  
Qy 526 GAGTCAATATATGAGAAATTTCTACTATCCAAATGGA---AGAAATCAACAGTAAGA 582  
Db 517 GAGTAAATGTGACACGACATTTCTATATTTCCAGAAACATCTGACAAACACAAAGA 576  
Qy 583 TGTTCAGAGAGTATTTACTGCTCAAAATGTAGTAAGAAAGTGTGAGGAGGACGCTTTAC 642  
Db 577 TGTTCACGATTAATCTGTTTACATGTTTGAAGCAATGCAAGGGGCGACGCTATA 636  
Qy 643 ACCAAAGATGCAATTCATTAAGTAAAGTAAAGACGACTCTATGAAAAAGATGTGATTT 702  
Db 637 GCACGACCATCTGACGACTCACAGACAGGCGTGTATGAAAGCAAAAGTCACTATTATC 696  
Qy 703 CTCCATCCCGCAGAGGAGAGGCTCTATATGTTTGGACCAACATGTTATCTATA 762  
Db 697 CCAAGAGATCCCAAGCTGCGCAAGAAATCATTTGTTTATGCAAAATCTTATATCTGTG 756  
Qy 763 GTTGAATTTCTGACAGAACAAACCAACAAAGAGCTCCAAAGCAAGCAAAATCAAAAA 822  
Db 757 ACTGAATTTTGTACTGAAAAAACACACAAATAAAGAGTCCAAACCTATAGAACAATAAG 816  
Qy 823 TGCATCTCCAGACACATGGGAAGTATCGTGTATCTGAGAGCTTTAAGAAAACT 882  
Db 817 TGCATCTCAGAAAGCAATGGGATGTAAATCATGAGCTCGAAGATTTTCAGATTATCT 876  
Qy 883 CCTATGAC---AACAGCGCACCAAAATCCCACTCTCTGCTGAGATGTGACAAAGA 939  
Db 877 CCAATGACGAATAATTAATTTACCTGCTCTACATTTTCTATGCTCAAGTCCAAACGCT 936  
Qy 940 ATTTGTGTGTGCTTGTGACAAATCTGGAACATGCGCACTGTAAACGCTCATCA 999  
Db 937 GTAGCTGTGTTGTACTTATTAATCTGGAACATGAATGAAGAGAGGAGCTTGTGGG 1056  
Qy 1000 CTGAATCAAGCAGGCGCTTTCTGCTGTCAGACAGTTGAGCTGGGCTGCTGGTGGG 1059  
Db 997 ATGAATCAAGCAGGAATTTGACTGATTTCAATTAATTAAGAAAGGAGCTTGTGGG 1056  
Qy 1060 ATGTGACATTTGACAGAGCTGCTCCATGTACAAAGTGAATCATACAGTAAACAGTGC 1119  
Db 1057 TTGGTCAATTTTGACAGTTTGTCTAAATTCAAAGTAACTCATAAATAATTTGATGT 1116  
Qy 1120 AATGACAGGACACATCTGCGCAAAAGATTAATGACGAGCTTCCAGGAGGAGCTCAATC 1179  
Db 1117 AACATTTACCAAAAGATCTCTGCAAACTGCTCAAGAAAGCTGATGTGTGCACTTCAAT 1176

Qy 1180 TGCAGCGGCTTCGATCGGCAATTTACTGTATAGAAAGAAATATC---CACTGATGGA 1236  
Db 1177 TGCAGGGGACTCAAAAGCAGGATTTTCAGCAATTTCCCAAGATTAACAGATCTTCGCT 1236  
Qy 1237 TCTGAATTTGTCTGACGAGATGGGAGCAACACATATAGTGGGTGCTTAAACAG 1296  
Db 1237 TCTGAATCATATTTACTAAGATGGGAAAGATTTATCAAAATAGCTTGTGTTGAGAG 1296  
Qy 1297 GTCAACAAAGTGTGCTCATCATCCACAGTCTGCTTTGGGGCCCTGACGCTCAAGAA 1356  
Db 1297 GTAAACAAAGTGGGACAGTATCCACACATGCTGTGGGGCCCTGCTGACAGAAAGAA 1356  
Qy 1357 CTAGAGAGCTGTCCAAATATCAGAGAGTTTACAGACATA----- 1398  
Db 1357 CTGAGAGCCCTGTCAAAATATGACAGATTACATAAGGACACGTATATCTGAAAGTTCA 1416  
Qy 1398 -----TGGTTAGATCAAGTTCAAGAACAT 1422  
Db 1417 TAAAGTCTGGGAAGTTTATCTTTTGAAGACATGCTTTTATGCCCATAAACATTAAT 1476  
Qy 1423 GGCCTCATGTATGCTTTTGGGGCCCTTTCATCAGAAATGAGAGCTGTCTCAGCGCTCC 1482  
Db 1477 GGCTTATTTGATGCTTTCAGACGAATTTCACTATGAAATGACAGATCTCAGAGAGCT 1536  
Qy 1483 ATCCAGCTTGAAGTAAAGGATTAACCTCCAGAAACAGCCAGTGTATGTCACAGTG 1542  
Db 1537 CTTCAGTTGGAAGTAAAGCTTTGAAATCCAGGCAAGAAATGATTAATGTATCACTG 1596  
Qy 1543 ATGCTGACACACCGCTGGGAAAGACACTTGTGTTTCTTACCTGAGACAGAGCT 1602  
Db 1597 CCTGTGATGATGACAGTAAAGATATATCTTCTTGTGTCACATGACGATACAAAG 1656  
Qy 1603 CCCCAGATCTTCTCTGAGATCCAGTGCAGAGCAAGAGTGGCTTGT-----A 1653  
Db 1657 CCAGCAATTAATTTCTCAAGATCCAAAGAAAAAATATCTACTCTCAGATTTTCAGAA 1716  
Qy 1654 GTGACAAAAACCAAAATATGCGCTACCTCCAAATCCAGGACATGTAAAGTTGGCACT 1713  
Db 1717 GGTGAATTAATTAATTCGGTCTGCGCTGCTTGTGAATCCAGGATATGTCAGACAGGCACT 1776  
Qy 1714 TGAATATCACTGTGCA-----GCAAGCTCACAACTGACCTGACTGTGACG 1764  
Db 1777 TGAATTTACAGCTTCGAAGACATATACCAATCTCAATTTGCTATCTGTGACATGACC 1836  
Qy 1765 TCCGCTGCTCCATGCTACCTGCTCCAAATTAACAGTACTTCCAAAGCAAGAGAC 1824  
Db 1837 ACTGAGCAAGAAAGCCCTACACACATCCCAATTAATGCAACTGTCCACATGACAAAT 1896  
Qy 1825 ACCAGCAATTTCCCAAGCCCTGTGTTATGCAATATTTGCCAAGAGGCTCCCA 1884  
Db 1897 ACAGCTCATTAACCTCAGCCAGTGTGTTTATGATGATGATGATGATGATGATGATGAT 1956  
Qy 1885 ATTCAGAGGCGAGTGTACAGCCCTGATGATGATGATGATGATGATGATGATGATGAT 1944  
Db 1957 GTTCTGGGAATCAATGTAACAGCCATTTAGAAATGAAAGGAGCAATCAAGTAACATG 2016  
Qy 2005 TATTTCAACAATATGACACGAATGTTGATATACAGTGTAAAGTGGGGCTCTGGAGGA 2064  
Db 2077 TATTTACAGATTTACATGGAATGTTGATGATGATGATGATGATGATGATGATGATGAT 2136  
Qy 2065 GTTAACGAGCAGAGGAGATGATACCCAGAGAGTGTGAGAGCTGTACATACCTGCGC 2124  
Db 2137 AAAACACAGCTAGGC-----TAAGTCAACACAGAAATTAAGCTCTATATGACGCGC 2190  
Qy 2125 TGGATGGAATGATGAATCAATGAATCAACCAAGAGCTGAAATTAATTAAGATGAT 2184  
Db 2191 TATGCTGAAATGGAATAATTAATGTAACCATTCCAAACTGAAAGTACAGATGATGTG 2250

QY 2185 GTTCAACACAGCAGAGTGTCTGAGCAGACATCTCGGAGGCTCATTTGTGCTCT 2244  
DB 2251 GAAGAGCTCAACACAGCAGCTTACGAGACTACCTCTGGAGGGTCTGTACTGTATCA 2310  
QY 2245 GATG---CCCAATGCTCCCATACCTGATCTCTCCACCTGGCCAAATCACCGACTG 2301  
DB 2311 GAGTGGCTCTCATATGATATCATCTTCAGGTGTCTGACCTGGTAAATTTAGACCTC 2370  
QY 2302 AAGCGGAAATTCACGGGGGCACTCTCATTTAATCTGACTTGAACAGCTCTCGGGATGAT 2361  
DB 2371 GAGGCTAAGTTTCAAGAG---ATCATATTCACCTTTCAGAGCTGCCCTGGCAAGTTC 2427  
QY 2362 TATGACCATGAGCAGAGCTCAGTATATCATGTAATGTAATGTAATGTAATGTAATG 2421  
DB 2428 CTGATTAAGAGAGAGAGTGTGAGCTATATTAAGATTAAGATTAAGATTAAGATTAAG 2487  
QY 2422 AGAGACAGTTCATGATGATCTCTCAAGTATGATGATGATGATGATGATGATGATGAT 2481  
DB 2488 CAGAGAGATTTTGTATTAAGCTGCTTAAATTAATCTCTGATGATGATGATGATGATGAT 2547  
QY 2482 AACTGTGAGAGCTCTTTTGTGTTAAACGAGAAACATTAATCTTTGAAATGCGACAGAT 2541  
DB 2548 GGTTCAGTGAAGTGTGATTTAAACGAGAACTCTTAATTAAGATGATGATGATGATGAT 2607  
QY 2542 CTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2601  
DB 2608 TTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2667  
QY 2602 GCACGAGTATCTTTGTTTATCTCTCCACAGACTCC 2636  
DB 2668 GCACAGCACTAATCTTATCTCTCCACAGAGACC 2702

RESULT 9  
AF001263 2820 bp mRNA 19-NOV-1997  
LOCUS Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.  
ACCESSION AF001263  
NID 92623766  
VERSION AF001263.1 GI:2623766  
KEYWORDS  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
REFERENCE 1 (bases 1 to 2820) Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
AUTHORS Elble, R.C., Widom, J., Goodwin, A. and Pauli, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
endothelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2820) Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
AUTHORS Elble, R.C., Widom, J., Goodwin, A. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES  
source location/Qualifiers  
1..2820  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone="3"  
/tissue\_type="lung"  
/cell\_type="aortic endothelial cells"  
194..2656  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86531.1"  
/db\_xref="pid:92623767"  
/db\_xref="gi:2623767"

CDS

Query Match 28.7%; Score 806; DB 3; Length 2820;  
Best Local Similarity 61.5%; Pred. No. 1e-203;  
Matches 1482; Conservative 0; Mismatches 880; Indels 48; Gaps 10;  
BASE COUNT 940 a 565 c 560 g 755 t  
ORIGIN  
/translation="MVICLVNLTFLILPEKSSMYLNNVGDVIAINPSVE  
DEKLNIKEMTASTYLFAHTRKRYFRANSLIPMTKRSSEYFIPKOSYQAD  
VIVANPYKYGDDPYTLQYRGCEGKATHEFPNLLTNNEITVSGKRVVEMAH  
RWGIDFENVDQPEYISKRNTLEATRGSTHIGINVKKCGGGCITSLCRDSOTG  
LYEACQTELPKKSQAKESIMPELHSTVECTEKTNTENAPNQNKNKSTWDY  
INNSFONTSPEEMNPPTHEPFLHSKROVCLVDKSGMSAEPLFOMNOAE  
LYLQIVIEGSLVGMVTPDSVAEIONHLTRIDNVVQKIKTAKLEOVANGTSCRG  
KAGFOALIHSDSTSGSEILITLDEDEINSCFEDVRSGLIHTIALGPSAAELE  
TUSNMTGGYREFNANDITGLINAFSRISSRSISITQOAIQLESKALKITGRKRVNGIV  
PYDSVNDPFEVFWTIOKPEIYLDPEKSKYKSDREKDLNTRSAIQPGIAET  
GWTYSLLNNHASSOMLTVTTRARSPDIPVIAATAHOSHTATVPSPMIYAVSQ  
GFLPYGISVIAIETEDGHOVTEILMNGACRDIVKNDIYSRETFDPEGRSLK  
VHAQARNMTARLNLNPOKNTLYVGVYENGKTIINPPPEKDLAKIETDSRLT  
SGSFVSGAPPNPNHPSYFPKSLITDLEAKKEDYIOLSWTAPGNVLDKGAASSFP  
MSRFHVQAKVLEQLQHOHOFQ"

QY 5 TCACAGGAGATGTACAGAGAGGGCCATTGAGAGTCTGTGTTCACTTGATCTTC 64  
DB 174 TTACTGTAACTATGCAAAATAGTCTCTGTGATGATGATGATGATGATGATGATGATGAT 223  
QY 65 ACCCTTCAGAGAGGGCCCTGAGTATTCATCTATTCAGTCAAGCAAAATGGCTATGAG 124  
DB 234 ATCTCTGGCTGG---AATGAAAGTTCAATGATGATGATGATGATGATGATGATGATGATG 290  
QY 125 GCATTTGCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 184  
DB 291 GCATTTGCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 350  
QY 185 AGGACATGGGAGCCGAGCATCTGTATCTGTTGATGATGATGATGATGATGATGATGATGAT 244  
DB 351 AGGAAATGGTAACTGAGGCTTCTTACTTCTGTTTATGACCAACAGAGGTTTATT 410  
QY 245 TCAAAATGTTGGCATTTGATTCCTGAAACATGAGAGAGAGGCTGATGATGAGAGC 304  
DB 411 TCAGGATGATGAGCATTTTATTCATGACCGTGGAAATCAAAATGATGATGATGATGATG 470  
QY 305 CAAATCTGAGACCTACAAAATGCTGATGTTCTGTTGCTGATGATGATGATGATGATGATG 364  
DB 471 CAAATCAAGATCATATGACCGGAGATGCTATGATGATGATGATGATGATGATGATGATG 530  
QY 365 ATGATGAGCCCTGACTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 424  
DB 531 GAGATGATCCCTATACACTTCAATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 590  
QY 425 TCACCTCCTGATTTCAATGAG 484  
DB 591 TTACTCCAAACTCTGTGACAAATATATTTCCATCATATGAGGCTCCGAGGAGAGGATAT 650  
QY 485 TTGTCCATGATGAGGCTCATCTACATGAGGAGGAGGATTTACAGATGATGATGATGATGAT 544  
DB 651 TTGTCCATGATGAGGCTCATCTCCCGCTGGGAGATTTATGATGATGATGATGATGATGATG 710  
QY 545 AATTCACCTATGCC---AATGAGAGAAATCAAGCAGATGATGATGATGATGATGATGATGAT 601  
DB 711 CATTCATATTTCCAG 770  
QY 602 GTACAAATGTAGTA---AAGAGGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658  
DB 771 GATATGATGAGGTTTCAAG 830  
QY 659 TCAATTAAGTACAG 718  
DB 831 GTGACTCAAG 890  
QY 719 CGAGAGAGGCTCTAATAGTTGACACATGTTGATGATGATGATGATGATGATGATGATGATG 778

Db 891 CTGCAAGGAATCCATTATGTTATGCCAGTCTCCATCTGTGATGAATTTTGTACAG 950  
Qy 779 AACAAACCCACCAAGAGAGCTCCAAACAGCAAAATCAATCAAAATATGCAATCTCCAGCA 838  
Db 951 AAAAAACACACATACAGAGCTCCAAACCTCAAAACAAATATGCAATGSCAAAGCA 1010  
Qy 839 CATGGGAAGTATCCGTGATCTGAGGACTTTAGAAAAACCATCTTATGACA-----A 892  
Db 1011 CATGGAGTATATCATGAACTCTGTGACTTTCAGAAATCATCTCCATGAGCAAGAAATGA 1070  
Qy 893 CACAGCCACCAATCCCACTCTCATGTCTCAGATGGAATGGAAGAAATGCTGTTAG 952  
Db 1071 ATCCACGACATCATCTCAATTTTCAATGCTCAAGTCCAAAGCGGGAGTCTGTTGG 1130  
Qy 953 TCCCTGACAAATCTGGAAGCATGCGAGCTGTACCGCTCAATCGACTGAATCAAGAG 1012  
Db 1131 TACTTGATTAATCTGGAAGCATGTCTGCAAGAACCGTCTCTTCAATGATCAAGCAG 1190  
Qy 1013 GCCAGCTTTCTGCTGCAGACAGTGTAGCTGGGCTCTGGTGGGATGTTGACATTTG 1072  
Db 1191 CAGAACTACTTGTGATTCAGATTATGAAAGGATCTTTAGTTGGATGTTACATTTG 1250  
Qy 1073 ACAGTGTGCGCCATGATGACAAAGTGAATCATGATTAAGATGAGTGGAGTGAAGGACA 1132  
Db 1251 ACAGTGTGCTGAATTCACAAATCATCTAACAGAAATTAAGTATGATTTTACCAA 1310  
Qy 1133 CACTGCGCCAAAGATTACCTGCAGACAGCTTCAGAGGAGCTCATCTGCAGCGGCTTC 1192  
Db 1311 AGATCAACGCCAAACCTGCTCAAGTATGATGTTGGAATCTTCAATTTGTAGAGGCTCA 1370  
Qy 1193 GATCGGCTTACTGTGATGAGGAAGAAATATC---CAACTATGATGATGAAATTTGTC 1249  
Db 1371 AAGCAGGATTCAGGCAAAATTAATCCAGAGTCAAGTCTCTGTTCTGGAATCATAC 1430  
Qy 1250 TCGTGCAGATGGGAGGAGCAACACTAATAGTGGGCTTTAACGAGTCAACAAAGT 1309  
Db 1431 TATTAACTGATGGGAGAGATTAATGAATTAATTAATCATGCTTTGAGATGTTAAACGAAGT 1490  
Qy 1310 GTGCCATCATCCACAGTGTGCTTGGGCGCTCTGCAGCTCAAGAACTAGAGAGCTGT 1369  
Db 1491 GTGCATCATCCACAGATGCTGTGGAGCCCTGCTGCCAAAGATGAGACATTTGT 1550  
Qy 1370 CCAAAATGACAGAGCTTTACAGACATATGCTTCAGATCAAGTTGCAACATGGCTCA 1429  
Db 1551 CAAATATGACAGAGATATC-----GTTTTTTTGCATTAAGCAATTAAGTACCTGCTTA 1604  
Qy 1430 TTGATGCTTTGGGCGCTTTTATCAGAAATGAGAGCTCTCTCAGCGCTCATACAGC 1489  
Db 1605 CTAAATGCTTTAGTGAATTTTATCTATGAAAGTGAAGCATCTACAGCGCTATTCAGT 1664  
Qy 1490 TTGAGATGAGGATTAACCTTCAGACAGCCAGTGTGAATGGCACATGATCTGG 1549  
Db 1665 TCGAAAGCAAGGCTTGAATAATACAGAAAGAAAGATTAACGCGACAGTCCGTAG 1724  
Qy 1550 ACAGCAACCTGGGAAGGACATTTCTTTCTTATCACTCGGCAAGCAGCGCTCCCAAA 1609  
Db 1725 ACAGTACAGTGGAAATGACATTTCTTTGTGTACATGAGAAATACAAAAACGAAA 1784  
Qy 1610 TCCCTCTGCGATCCCATGAGACAGACA-----AGTGGCTTTGATGAGACA 1660  
Db 1785 TTGTTTCCAAAGTCCAAAGAAAGAAATATTAACCTCGGATTTCAAGAAAGATTAAGT 1844  
Qy 1661 AAAACCAAAATGGCTACTCTCAATCCAGCATGTCTAAGTTGGTCACTTGGAAAT 1720  
Db 1845 TAAATATGATCTGCTGCTCAATCTGATTTGAGAGACAGTATGTTGACTT 1904  
Qy 1721 ACAGTCT-----GCAAGCAAGCTCAAAACCTTGACCCGAGTGTACAGTCCCGTG 1771  
Db 1905 ACAGCTTTCTAATATATCATGCGAGCTCTCAAAATGCTTAACAGTACAGTACACTCGAG 1964  
Qy 1772 CGTCAATGCTACCTGCTGCTCAATACAGTGTACTTCAAAACGAACAGACACAGCA 1831  
Db 1965 CAAGAGTCTCTATATACCCCGAGTATGTCAGACAGCTCAAGATCAACATACAGCAC 2024

Qy 1832 AATTCGCCAGCCTCTGTGATGTTATGCAAAATATTGCCAAGAGCCTCCCAATTTCTCA 1891  
Db 2025 ATATCTAGCCCAATGATGTTATGTCACAAAGTCAAGAGGTTTTCCTGCTACTGG 2084  
Qy 1892 GGGCAGTGTACAGCCCTGATTTGAATCAGTGAATGGAATAAAGATTACCTTGGAACTAC 1951  
Db 2085 GAATCAGTGAATGCGCATTAATGAAACCGAGATGAGATCAAGTAAATATGAGTGTCT 2144  
Qy 1952 TGGATTAATGAGCAGTGTCTGATGCTATAGATGAGCGGTGTCTACTCAAGATTTTCA 2011  
Db 2145 GGGCAATGTTGAGTGTGTGATGCTATGTCAGAAATGATGGATCTACTCAAGATTTTCA 2204  
Qy 2012 CACTTATGACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2071  
Db 2205 CAGATTAATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2264  
Qy 2072 CAGCAGCGAGAGTGTATACCCGAGAGTGGAGCAGTGTACATGCTGCTGATGATG 2131  
Db 2265 CGGCTAGGCTTAATTAATTAAGACACACACAGAAAGTTCTATGTTCCAGGCTACGTTG 2324  
Qy 2132 AGAATGATGAATATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2191  
Db 2325 AAAAGGTAAATATATCTGTAACCCAGACCTGAAGTCAAGATGATGATGATGATGATG 2384  
Qy 2192 ACAAGCAAGTGTGTTACGACGAACATCTTGGGAGGCTCATTTGTGCTTGTATGTC 2251  
Db 2385 CTAAATGAAAGATCTTACGAGTACCTTGGAGGCTCTTTACTGTATCAGAGCTC 2444  
Qy 2252 CAAATGCTCCCA---TACCTATCTCTCCACCTGGCCAAATCAGCAGCTGAAGCGG 2308  
Db 2445 CTCCTCCGTGATACACCTCTGTTGCCACAGTAATTAATGATGATGATGATGATGATG 2504  
Qy 2309 AAATTCAGGGGGGAGTCTATTAATCTGACTTGACAGCTCTTGGGAGTATATGACC 2368  
Db 2505 AGTTCAAAGAG---ATTAATATCAACTTGTATGAGACAGCCCTGGCAATGCTTAGTA 2561  
Qy 2369 ATGCAACAGC 2378  
Db 2562 AAGGAAAAGC 2571

RESULT 10  
AF001262  
LOCUS AF001262 3288 bp mRNA MAM 19-NOV-1997  
DEFINITION Bos taurus clone 2 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
complete cds.  
ACCESSION AF001262  
NID 92623764  
VERSION AF001262.1 GI:2623764  
KEYWORDS  
SOURCE  
ORGANISM  
Bos taurus.  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE  
1 (bases 1 to 3288) Gruber A.D., Abdel-Ghany M., Levine R.,  
Eblle R.C., Widom V., Goodwin A. and Pauli B.U.  
Cloning and characterization of Lu-ECAM-1 suggest it is an  
endothelial chloride channel  
JOURNAL  
REFERENCE  
2 (bases 1 to 3288)  
AUTHORS  
Goodwin A., Gruber A.D., Abdel-Ghany M., Levine R.,  
Eblle R.C., Widom V., and Pauli B.U.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES  
location/qualifiers  
1..3288  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"

CDs /clone="2"  
/tissue\_type="lung"  
59..2443  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AA086530.1"  
/db\_xref="pid:g2623765"  
/db\_xref="gi:2623765"

BASE COUNT 1043 a 707 c 645 g 893 t  
ORIGIN  
Query Match 28.5%; Score 800.6; DB 3; Length 3288;  
Best Local Similarity 61.9%; Pred. No. 2.7e-202;  
Matches 1442; Conservative 0; Mismatches 844; Indels 45; Gaps 9;  
/translation="MYCLNVLFLTLHPKMSMVLINNGYDGIATVIAINPSYE  
DEKLENIEMVETASTYLFHAKRVRVNSILPMKSKSEFIKROSYODAD  
VIVANDYLKGDPTLOYRGKRGKXVHFTNPLLNTHNHYSGRIVFHEAHL  
RMGIDFVNDPFIYSKNTIEATRCSTHITGINVYFKCGGSCITSLCRDSOTG  
LEACCTFLPKSQTAKESIMFPLSHVTEFCTEHTNEAPNLONKCNKSTWDV  
INNSVDFONTSPTEENPPTHEFLSLKSKORVCLVLDKSSMSDELFPQNOAE  
LYLIVIEKSGLVGMTEDVSAIEONHLRIDDDNYQKRTAKLPVANGSTICRGL  
KGFQAIHSDOSTGSEIILTDGENDINSCFEDVKRSGAIHITIALGPSAKELE  
TISNMGYREFRANNDIGLTNAFSISRSISGSIQOATOLESKLAKIRGRVNGVT  
PVDVNGDFFVVTIOTKPRIVADPDKRYKTSDFEDLNTIRSAILOJPGIET  
GTWTYSLNNHASSOMLIVYVITRASPPIPVIAHNSHTAHPSPIVAYQVSQ  
GFLPVAGISVIAIIEDEGHQVLELMDGAGRDVVKNDGIISRTFYDYGGRSFLK  
VHAQRNNTARLNROPKNKVLVYDGYENGRILINPPEYKDLAKKIDFESFLT  
SGSFVSQAPPPGNHPSVFPSPKIIDLAKKEKEDIQLSTAPGNVLDKGES"

QY 83 TGAGTAATTCACATTCACAGTCAACACATGGCTATGAGGCAATGTTGTTGCAATCG 142  
DB 114 TGAAGATTCAAGTGAATTTGATTAACAATGGGATGATGGCATTTGATTCGCAATTA 173  
QY 143 ACCCATATGCGCAGAGATGAACACTCATTCACAAATAAAGAGCATGGTACCAGG 202  
DB 174 ACCCGATGTCGAGAGATGAACAACTCATTAACAAATGAAGATGTAACTGAAG 233  
QY 203 CATCTCTGATCTGTTGAAGCTACAGAAAGCATTTTCAAAATGTTGCCATTT 262  
DB 234 CTCTACTTACCTGTTTCAATGCAACCAAGAGATTTTTCAGGAATGTAGACATTT 293  
QY 263 TGATTCCTGAACATGGAACAGCAAGGCTGACTATGTGAGACCAAACTTGACACTACA 322  
DB 294 TAATTCATATGACCTGGAATCAAAATCTGAGTACTTACACAAAACAAGATTCATATG 353  
QY 323 AAAATGCTATGTTGTTGCTGAGTCTACTCTCCAGTAAATGATGAACCTTACACTG 382  
DB 354 ACCAGGAGATGTCATATGTTGTAATCCCTATTAATAATGAGAGATGCCCTATACAC 413  
QY 383 AGCAGATGGCACTGTGAGAGAGAGGAGGATCCACTCACTCTGATTTCAATG 442  
DB 414 TTCAATATGGAAGCTGTGAGAAAAGGAAATATATACATTTTACTCCAACTTCTGT 473  
QY 443 CAGGAAAAAGTTAGCTGAATATGACACACAGAGTGGGCAATTTGTCATGAGTGGCTC 502  
DB 474 TGACTATATATTTCCACATCTATGAGGTCGCGAGGAGAGATTTGTCATGAGTGGCCC 533  
QY 503 ATCTAGATGGGAGATTTGACGAGTACATATGATGTAAGTAATCTATTAATCC--A 559  
DB 534 ATCTCCGCTGGGAAATATTTGATGAGTATATATGTGACCGCAATCTATATTTCCGAA 593  
QY 560 ATGGAAGATACAGAGTAAAGATGTTACAGAGTATTACTGTACAAATGAG--TAA 616  
DB 594 AGAAGACTATGAGCAACAAGATGTTCAACTATATGATTAATGATGGTTTCA 653  
QY 617 AGAAGTGTACAGGAGGAGCTGTTACACCAAAAGATGACATTCATTAAGTAAACGAG 676  
DB 654 AGAAGTGTCCCTGAGGAGAGCTGTATTAACAGTCTATGACACAGTGTACTACAGAGG 713  
QY 677 TCTATATAAAGAGTGTGAGTTGTTCTCCATCCGCGCAGAGGAGAGGCTTCTATAA 736

DB 714 TGTATGACAAAATGTACATTCCTTCCAAAAAATCCAGACTGCAGAGATTCATTA 773  
QY 737 TGTGTGCACAACTGTGATTTCTATAGTGAATTCGTACAGAACAAACACACAAAG 796  
DB 774 TGTATATGCAAGTCTTCAATCTGTGACTGAATTTTGTACAGAAAAACACAAATAC 833  
QY 797 AACCTCCAAACAAAGCAAAATTCAAAATTCGAAATCTCCGAAAGCAGATGGAAATGATCC 856  
DB 834 AACCTCCAAACCTACAAAACAAAAATGTGCAATGCAAAAAGCAGATGGATGATATGA 893  
QY 857 ATCTGAGGACTTTAAGAAAACACCTCATATGACA-----ACACAGCCACCAATCCCA 910  
DB 894 ACCTCTGTGACTTTCAGAAATACATCTCCCATGACAGAAAATGAAATCCACGACTCATCT 953  
QY 911 CCTTCTCATTTGCTGAGATTTGACAAAAGATTTGTTAGTCTTGAACAACTGGA 970  
DB 954 CATTTTCATTTGCTCAAGTCCAAACACGCGGATGATGTTGTTGATTTGAATTAATCTG 1013  
QY 971 GCATGGCAGCTGTAACCGCTCAATCGACTGATATACAGAGCCAGCTTTCTCTCTG 1030  
DB 1014 GCATGCTGCAGAGAGACGCTCTTTCAAATGATCAACAGCAGCACTATCTGATTC 1073  
QY 1031 AGACAGTTGAGTGGGCTGCTGGGTTGGGATGGTGCATTTGACAGTGGCCCATGTG 1090  
DB 1074 AAGTTATGAAAAGGATCTTTAGTTGGATGGTTACATTTGACAGTGTCTGAAATCC 1133  
QY 1091 AAGTGAAGTCAATACATTAACAGATGAGGAGTGCAGAGGACACACTGCCAAAAGATTAC 1150  
DB 1134 AAATATCATCTAACAAATAATACATGATATATGTTAACAAAAGATCACCGCAAACTGC 1193  
QY 1151 CTCGACAGCTTCAGAGAGGAGAGCTCATCTGCAGCGGCTTGATGGCATTTACTGTGA 1210  
DB 1194 CTCGAATGACTAATGATGAGAACTCAATTTGTAGAGGCTCAAAAGAGATTCACAGCAA 1253  
QY 1211 TTGAGAGAA--AATATCAACTGATGATGATGAAATTTGCTGAGAGATGGGGAAG 1267  
DB 1254 TTATCCAGATGACAGAGATCTCTGTTGTTGATTAATCATATTAATGATGAGGGAAG 1313  
QY 1268 ACAACATTAATGAGTGGTCTTTAACGAGGTCAACAAAGTGTGCTATTCACACAG 1327  
DB 1314 ATATGAATAATTAATCATCTTTGAGAGATGTAAAGAGATGGTGAATCATCACACCA 1373  
QY 1328 TCGCTTTGGGCGCTCTGACGCTCAAGAACTAGAGAGCTGTCCAAAATGACAGAGGTT 1387  
DB 1374 TTGCTCTGGAGCCCTCTGCTGCCAAGAACTGAGCAATGTCAAATATGACAGGAGAT 1433  
QY 1388 TACAGACATATGCTGATGATCAATGATGACAGAACTGAGCAATGATGATGCTTTGGGGCCC 1447  
DB 1434 ATC-----GTTTTTTGCCAATTAAGACATTAATGCTTACTTAATGCTTTTCAGTAA 1487  
QY 1448 TTTCATCAGAAATGAGGCTGTCTCTACAGGCTCATTCAGCTTGAAGATTAAGGATTA 1507  
DB 1488 TTTCATCTGAATGAGAGATCATCTCAGAGGCTATTCAGTGTGAAGCAAAAGCTTGA 1547  
QY 1508 CCTCCAGAAAGCCAGTGGATGATGACAGAGTATCTGAGACAGACACCGTGGGAAAG 1567  
DB 1548 AAATTCAGAGAAAGAAAAGTAAAGGACAGAGTGTCTAGACATTAAGTGTGAAATG 1607  
QY 1568 ACATTTGTTCTTATACCTGAGCAACGAGCTCCCAAAATCTCTGAGATCCCA 1627  
DB 1608 ACATTTCTTTGTTGATGATGACAGATACAAAACCAAAATTTGTTCTCCAAAGATCCA 1667  
QY 1628 GTGAGCAGAGCA-----AGGTGGCTTTGATGACAAAACACCAAAATGGCCT 1678  
DB 1668 AAGGAAAGAAATTAACCTCGGATTTCAAGAAAGATTAAGTAAATTTGATGATGCTC 1727  
QY 1679 ACCTCCAAATCCAGGCAATTTGTAAGTGTGCACTGGAATATACATCT-----GC 1729  
DB 1728 GTCTGCAAAATACCTGATTTGAGAGAGAGTACTTGACCTTACACGCTTCAAAATATC 1787  
QY 1730 AAGCAGCTCACAAAACCTTGACCCCTGATGACAGTCCGCTGCTCATGCTACCTG 1789  
DB 1788 ATCCAGCTCTCAAAATGCTAAACAGTACAGTACCACTGAGACAGAAATCTTACTATAC 1847



## JOURNAL

Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases. Reiko Itoh, Institute for Molecular and Cellular Biology, Osaka University, Yamada-oka 1-3, Suita, Osaka 565-0871, Japan (E-mail: reiko@imcb.osaka-u.ac.jp, Tel: 81-6-6879-7992, Fax: 81-6-6877-1922)

## FEATURES

Location/Qualifiers

source 1. 3604  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="lambda 13802"  
/clone\_1ib="lambda zap"  
/dev\_stage="adult"  
/map="1p32"  
/tissue\_type="corneal epithelium"  
42..2873  
/codon\_start=1  
/product="chloride channel protein"  
/protein\_id="BAA77810.1"  
/db\_xref="PID:d1041578"  
/db\_xref="PID:94887601"  
/db\_xref="GI:4887601"  
/translation="MTQSIAPICNLKFTLLVALSELPELGAQVQLQDNGYGLLAINPQVENQNLISNIKEMTEASFLFNATKRRVFEFNILIPATNRKANSKIK OESYEKANVITDMYGAGDDPYTLOYRGCQEGKYIHFTPNELNDLRNAGSRG VEFHEMAHLRMGVEDEYINDKPFYINGMOIKVTRCSDSDITGIFVCEGSPQENCII SKLREGCTFIYNSQNTKASIMPMOSLSYVEFCNASHNEARLNQKQSLNSAM DEYTDADPHSHFPMNGTELPPTPFSLVQADKVCVLDVSSKMAEDRLQLQQA AEFLYMOIVEIHFTFVGIASFDSKEIRALQHINSNDKLLVSLPTVSAKTDISI CSGLKRGFEVLEKNGKAKAYSVMLVTSDDDLKNCLEPTVSGSTIHSIALGSSSA PLEELSLRTGLKFEFVPDISNSNIMDFSRISSTGDTIFQOHLQESTGYNVPH OUKNTVDYDNTGNDMLTVMQASGPPIILDDGRRKYVNNFTINTPTFASLIM BETAPRGHNTYTLNNTNHSIALKATVYSRASNSAVPPATVATPYERDSLHPHYMI IANVQGFPIINATVATVEPETDPTLRLLDQAGADVINKNGITSRYEFSRAAN GRYSLSKGVNHSPSISTPAHSIPGSHAMVPGYTAGNIOANAPRKSVAGRNEERKQ FSRVSGGSFVLYGVAPGHPVPKIIDLEAVKVEBELISWTAPGEDDQCAQ STEIRMSKSLQIODFNNAILVNTSKRPQOAGIREITFSPOJSTNGPEHONGET HESHRIYVAIRAMDNRNLSOASVNTAIOAPLFIIPPSDVPARDYLLINGVLAMLLIG IICLIIVTTHHLSKRKRDRKENGKTL"

## CDS

BASE COUNT 1113 a 740 c 732 g 1017 t 2 others

## ORIGIN

Query Match 19.6% Score 550.6; DB 9; Length 3604;

Best Local Similarity 55.4% Pred. No. 7.4e-136; Mismatches 1039; Indels 66; Gaps 13;

Matches 1370; Conservative 0;

98 TTCAGCTGAACAACATGGCTATGAAGGCATGTGCTGCAATCGACCCCAATGTGCAG 157  
139 TACAGCTTCAGACATGGGTATATGATGCTCATTCGCATTAACTCCAGACTACCTG 198  
158 AAGATGAAACACTCATTTCAACAAATAAGACATGGTGAACCCAGGACATCTGTATCTGT 217  
199 AGAATCAGAACCTCATCTCAACATTAAGAAATGATTAAGCTTCAATTACTTACTTAT 258  
218 TTGAAGCTACAGAAAGCATTTTATTCAAAAATGTGGCATTGTTGATCCCGAAGCAT 277  
259 TTAATGCTACCAAGAGAAATATTTTTCAGAAATATAAGATTTTAATACCTGCCACAT 318  
278 GGAAGACAAGAGCTGACTATGTGAGACAAAATTGAGACCTCAAAAATAGTGATGTTG 337  
319 GGAAGCTAA---TAATAACAGCAAAATAAACAAGATCATATGAAAGGCAAAATGTCA 375  
338 TGGTGTGAGTCTACTCTCCAGGTAATGAACCTACACTGAGCAGATGGGCACT 397  
376 TAGTGACTAGCTGTATGTGGGCGACATGAGATGATCCATACACCTCAATACAGAGG 435  
398 GTGGAAGAAAGGTGAAGAGATCCACCTCACTGATTTTCAATTTGAGAGAAAAATT-- 456  
436 GTGGAAAGAGGGAATAATTCATTATTCACACCTAATTTCTACTGAATGATTAACCTTAA 495  
456 -AGCTGAATATGGACCAAGATAGGCAATTTGTCATGATGGGCTCATCTCAATGAGTGG 514  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

496 CAGCTGGTACGGATCAGAGGCGGAGTGTGTTGTCATGATGGGCCACCTCGTTGGG 555  
515 GAGTATTTACAGATCAATTAATGATGAGAAATTTACTTATTCATGAGAA---GATAC 571  
556 GTGTGTGATGATGATATCAATGACAAACCTTGTACTTAATGGGCAAAATCAATTA 615  
572 AAGCATATGATGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 631  
616 AAGTGAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666  
632 GCAGCTGTACACCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 691  
667 GTCTTGGCCCCCAAGAAACTGTATATATAGTAG-----CTTTTAAAGAAAGAT 717  
692 GTGAGTTTGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 751  
718 GCACCTTATCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777  
752 TGTATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 811  
778 TATCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837  
812 AAAATCAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 871  
838 AGAACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897  
872 AGAAAACACTCTAT-----GACACACAGCCACCAAAATCCACCTTCTCATCTGCTGC 925  
898 ACCACAGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 957  
926 AGATTGGAACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 985  
958 AGGCTGGCAGAAAGTGTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1017  
986 ACCGCTCATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1045  
1018 ACAGACTCTTCACTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1077  
1046 GGTCTGTGGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1105  
1078 ATACCTTCTGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1137  
1106 AGATTAACAGTGGAGTACAGGACACACCTGCCCAAAAGATTAC-----CTGCAGCAG 1159  
1138 AATTAACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1197  
1160 CTTACAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1218  
1198 CTAAACACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1257  
1218 -GAATATCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1276  
1258 TGAATGAAGAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317  
1277 TAAGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1336  
1318 TTGGCAATGCTTACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1377  
1337 GGCCTCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1396  
1378 GTTACATGTCAGCCCCCAATCTGAGAGATTTATCAAGCTTTACAGAGAGTTTAAAGTTCT 1437  
1397 ATGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1456  
1438 TTGTTCAAGATATATCAAACTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1497  
1457 GAAATGAGAGCTGCTCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1516  
1498 GAATCGAGACATTTTCCAGCAATATTCAGCTTGAAGATGAGTGAAGATGCAAAAC 1557  
1517 ACAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1576  
1558 CTCACATCAATTTGAAAAACACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1617



```

OY 1577 TTCTTATCAGCTG---GACACAGCAGCCTCCCAATCTTCTCTGGGATCCAGTGGAC 1633
    ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1618 TTCTAGTTACGTGGCAGGCCAGCTGTCTCTCTGAGATTATTTAGTACCTGATGCT 1677
OY 1634 AGAACA-----AGTGGCTTTGTAGTGACAAAACCCAAATATGGCTTACTCCAAA 1687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1678 GAAATACTACACAATATATTTTATACCAATCTAACTTTTGGAGAGTACTGTTTGA 1737
OY 1688 TCCCGAGCATCTTAAGTTGGCACTTGGAAATACAGTCTGCAAGAGCT----- 1739
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1738 TTCTGAGAACACCTTAAGCTGGGCACTGACTTACACCTGAAACAAATACCATCTCTC 1797
OY 1739 CACAAACCTTGAACCTGACTGTACGCTCCGCTGCTCAAGTCAATGCTACCTGCTCCATTA 1798
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1798 TCGAAGCCCTGAAAGTACAGTACCTCTCGGCTCCAACTCACTGCTGCCACAGTTGAGC 1857
OY 1799 CAGTGACTTCCAAACGACACAGACACAGCAATTCGCCAGCTCTGGTGGTTATG 1858
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1858 CTGTGGAAGCCTTTGTGGAAGAGACAGCCTTCATTTCTCATCTGTGATGATTATG 1917
OY 1859 CAAATATTCGCAAGAGCCTCCCAATTCAGAGGCCAGTGTCAAGCCTGATTGAT 1918
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1918 CCAATGTGAACAGGAGATTATTCATCTTAATGCTCACTGTCCACAGTTGAGC 1977
OY 1919 CAGTGAATGGAACAGATTACTTGGAACTACTGATATGAGAGAGTGTGATGCTA 1978
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1978 CAGAGACTGAGATCCTGTTACCTGAGACTCTCTGATGATGAGCAGAGTGTGATGTTA 2037
OY 1979 CTAAAGATGAGGCTGTCTACTCAAGTATTTTCAACATTATGACAGCAATGTGATACA 2038
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2038 TAAATAATGTGAATTTTACTGAGGATTTTCTCTTGTGTCGAATGTGATATATA 2097
OY 2039 GTGTAAATGCGGGCTCTGGAGAGATTACGACAGCCAGACGAGAGATACCCAGC 2098
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2098 GCTTGAAGTGCATGTCATCTCTCCAGCATAGCAGCCAGCCAGCTCATTTCCAG 2157
OY 2099 AGATGAGACAGCTGATACCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2158 GGAATCTATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2217
OY 2159 CAGACCTGGAATTAATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2218 CAGGAATCAGTAGGACGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2274
OY 2219 CCTGGGAGGCTCATTTGTGCTCTGATGTCCTGATGTCCTGATGATGATGATGATGATGATGATGAT 2278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2275 GCTGAGGAGGCTCTCTTTCATGCTGCGGAGTTCCAGCTGGCCGCCAGCTGATGTTTC 2334
OY 2279 CACCTGCGCAATACACGACCTGAAGGGAATTCACGGGGGAGTCTCATTAATCTGA 2338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2335 CACCATGCAAAATTAATTAAGCTGGAAGCTGTA--AAGTGAAGAGGAATGACCTAT 2391
OY 2339 CTGGAACAGCTCTGAGGATGATTAATGACATGGAACAGCTCACAGATATCATTTGAA 2398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2392 CTGGAACAGCTCTGGAAGACTTTGATCAGGCGGAGGCTGACAGCTATGAATAGAA 2451
OY 2399 TAAGACAGATATCTTATCTCAGAGACAAGTTCAATGAATCTCTTCAAGTGAATCTA 2458
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2452 TGAAGTAAGTCTACAGATATCCAAAGATGAATTAACAAATGATTTTAAATATACAT 2511
OY 2459 CTGCTCTCATCCAAAGAGCAACCTGAGGAAGTCTTTTGTTTAAACAGAGAAACA 2518
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2512 CAAAGGAATCCTCAGCAAGCTGCATCAGGAGATATTAAGCTTCCACCCCAATTT 2571
OY 2519 TTACTTTTGAATG 2533
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2572 CCACGAATGAGCTG 2586
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 13  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS 5

DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hclcal1) gene, complete cds.
ACCESSION	AF039401
NID	94009459
VERSION	AF039401.1 GI:4009459
KEYWORDS	human.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 35278) Gruber A.D., Elble R.C., Ji H.L., Schreier K.D., Fuller C.M. and Pauli B.U.
TITLE	Genomic cloning, molecular characterization, and functional analysis of human hclcal1, the first human member of the family of Ca <sup>2+</sup> -activated Cl <sup>-</sup> channel proteins
JOURNAL	Genomics 54 (2), 200-214 (1998)
MEDLINE	99047526
REFERENCE	2 (bases 1 to 35278) Gruber A.D., Elble R. and Pauli B.U.
AUTHORS	Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
JOURNAL	Location/Qualifiers
FEATURES	1..35278
source	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="1"
	/map="1p22-p31"
	1..1617
	/gene="hclcal1"
	1..33522
	/gene="hclcal1"
	767..1285
	/rptc_family="L1 transposable element"
	/rptc_type="dispersed"
	1587..1596
	/gene="hclcal1"
	/note="putative"
	1618..1708
	/gene="hclcal1"
	/number=1
	join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)
	/gene="hclcal1"
	/product="calcium-dependent chloride channel-1"
	1962..2383
	/gene="hclcal1"
	/number=2
	join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)
	/gene="hclcal1"
	/codon_start=1
	/product="calcium-dependent chloride channel-1"
	/protein_id="AAC95429.1"
	/db_xref="PIR:94009460"
	/db_xref="GI:4009460"
	/translation="MGPFKSSVFLILHLLEGALNSLIQNNNGYEGIVAVIDNPV EDTLIQIKDNTQASLYLFEAKGKRFYKNAVLIIPETWKTKADVRKLETTYNA DVLVAESTPQNDPEPTDQNGCKEKEERHLIPDPAKKLAEGPOGAEVHEAH LRMGFEEDYNDEKFLYLSNGRIQAVRSAGITGNVYKCGGSCCTKCTRETKVGL YKGECEFLYRSQTEKASIMEAHDVSIIVECTEONHNKAPKQNKCLRSTWEVI RDSDEFKTTDPNTPQNPFTSLQIGORIVCLVLDKSGMATGNRLRNQGLFL LQTVETGSMVAVTFEDSAHYOSELQINSQSDRLAKRLPAASGSLTICGSLSA FTVIRAKRPTDGSSEIVLITDGEANTISGCFEYVOSGAILHTVALGSAQDELSK MTGQLQITASDVONKCIIDAFGLAISGNGNAGVRSIOLESKGLTIONSOMAGTVL DSYVGKDTLFLITWTPQILMDPSQKQSGVYDKNTKMAVLIOIIGLAKGVTKY SLQASSQTLITLVISRASTNLPPIVYSKTKNTKTSFSPSLVYVNAIRGASPIISA SVTALIESVNGKVTLELLDNGAGADAKIDGVSRYFTTYIDNGRISVAVRALGCVN



```

AARRVTPQSGALYIPGMIENDEIOMNPBPINKDDYQHKOVCESPRSGGSFVAS
DVPNAPIDPLFPFGOITIDNANIEHSGSLNLTWTRAPGDYDGTAKRIIRISTSLD
LNDKFNESLQVNTALIPKANSSEVFLKPEPNTIFENSTDLFIALQAVDKVLDSEI
SNIAVRSLEIPPOTPPEPSPBETSAPCNHINSITPQIHILIKMKWIGLOLSIA
"
exon      6667..6807
           /gene="hclclal"
           /number=3
exon      6979..7126
           /gene="hclclal"
           /number=4
exon      9696..9801
           /gene="hclclal"
           /number=5
exon      15454..15631
           /gene="hclclal"
           /number=6
exon      18591..18809
           /gene="hclclal"
           /number=7
exon      19783..20010
           /gene="hclclal"
           /number=8
exon      22253..22427
           /gene="hclclal"
           /number=9
exon      24522..24628
           /gene="hclclal"
           /number=10
exon      26644..26859
           /gene="hclclal"
           /number=11
exon      27447..27708
           /gene="hclclal"
           /number=12
exon      28766..28936
           /gene="hclclal"
           /number=13
exon      31840..32079
           /gene="hclclal"
           /number=14
exon      32919..33522
           /gene="hclclal"
           /number=15
BASE COUNT 11245 a 6870 c 7230 g 9933 t
ORIGIN
Query Match      15.5%; Score 435.4; DB 11; Length 35278;
Best Local Similarity 99.8%; Pred. No. 3.1e-105;
Matches 436; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2376 AGCTCAGTATATATCATTCGATAGTACAGTATATTCGTGACGACAAAGTTCAA 2435
    |||||||
DB 32917 AGCTCAGTATATATCATTCGATAGTACAGTATATTCGTGACGACAAAGTTCAA 32976
QY 2436 TGAATCTCTCAAGTAACTACTGCTCTCATCCCAAGAAAGCAACTCTGAGGAAGT 2495
    |||||||
DB 32977 TGAATCTCTCAAGTAACTACTGCTCTCATCCCAAGAAAGCAACTCTGAGGAAGT 33036
QY 2496 CTTTGTGTTTAAACGAGAAACATCTTGTGAAATGGACAGACTTTTCAATGCTAT 2555
    |||||||
DB 33037 CTTTGTGTTTAAACGAGAAACATCTTGTGAAATGGACAGACTTTTCAATGCTAT 33096
QY 2556 TCAGGCTGTGATAGTGCATCGAATCAGAAATATCCAAATGACAGAGTATCTTT 2615
    |||||||
DB 33097 TCAGGCTGTGATAGTGCATCGAATCAGAAATATCCAAATGACAGAGTATCTTT 33156
QY 2616 GTTATTCCTCCACAGACTCCGACAGACACTAGTCTGATGAAACGTCTGCTCTTG 2675
    |||||||
DB 33157 GTTATTCCTCCACAGACTCCGACAGACACTAGTCTGATGAAACGTCTGCTCTTG 33216
QY 2676 TCCTAATATTCATATCAACAGACCAATCTCTGCATTCACATTTTAAAAATATGTGGAA 2735
```

```

|||||
DB 33217 TCCTAATATTCATATCAACAGACCAATCTCTGCATTCACATTTTAAAAATATGTGGAA 33276
QY 2736 GTGATAGGAGAACTGCAGCTGTCAATAGCCTAGAGGCTGAATTTTGTGAGATAAATAA 2795
    |||||||
DB 33277 GTGATAGGAGAACTGCAGCTGTCAATAGCCTAGAGGCTGAATTTTGTGAGATAAATAA 33336
QY 2796 ATAAATCATTCATCTT 2812
    |||||||
DB 33337 ATAAATCATTCATCTT 33353

RESULT 14
AF001264      1265 bp      mRNA      MAM      19-NOV-1997
LOCUS      Bos taurus clone 4 endothelial adhesion molecule Lu-ECAM-1 mRNA,
DEFINITION      complete cds.
ACCESSION      AF001264
VERSION      92623768
KEYWORDS      AF001264.1 GI:2623768
SOURCE      Bos taurus.
ORGANISM      Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovineae; Bos.
REFERENCE      1 (bases 1 to 1265)
AUTHORS      Eblie,R.C., Widom,J., Gruber,A.D., Abdel-Ghany,M., Levine,R.,
              Goodwin,A. and Pauli,B.U.
TITLE      Cloning and characterization of Lu-ECAM-1 suggest it is an
              endothelial chloride channel
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 1265)
AUTHORS      Eblie,R.C., Widom,J., Gruber,A.D., Abdel-Ghany,M., Levine,R.,
              Goodwin,A. and Pauli,B.U.
TITLE      Direct Submission
JOURNAL      Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,
              Ithaca, NY 14853, USA
FEATURES
     source
         1..1265
             /organism="Bos taurus"
             /db_xref="taxon:9913"
             /tissue_type="lung"
             /cell_type="aortic endothelial cells"
             /clone="4"
             109..1137
                 /note="endothelial adhesion molecule; chloride channel;
                 secreted form"
                 /codon_start=1
                 /product="Lu-ECAM-1"
                 /protein_id="AAB86532.1"
                 /db_xref="PID:92623769"
                 /db_xref="GI:2623769"
                 /translation="ATVLCINVTIFLHLBQKSMVNLINNGDGIIVAINPSVE
                 DEKLIENTKEMTASTYLEFHRTRKRVYFRANVSLIPMTWKSSTYFIPKDSYDQAD
                 VIVANPYLKGDPTLQYGRGEGKXITHEFPNLLTNNHITVSGRVEVYHVAHL
                 RWGIDEVNVDQPFYISRNKTIETARCSITHIIGINVVRKCPGGSCITSLCRDSOTG
                 LYEAKTCELPKRSOTAKESIMEMPSLHSTVECTEKTNTETAPNLQNMKNCKSTWDY
                 IKNSDPQNTSPMTEMNPTHTPFTSLKSKQVQVLVDKSGMSAEDIYLLALIKI
                 FKLIGNIT"
CDS
     109..1137
         /note="endothelial adhesion molecule; chloride channel;
         secreted form"
         /codon_start=1
         /product="Lu-ECAM-1"
         /protein_id="AAB86532.1"
         /db_xref="PID:92623769"
         /db_xref="GI:2623769"
         /translation="ATVLCINVTIFLHLBQKSMVNLINNGDGIIVAINPSVE
         DEKLIENTKEMTASTYLEFHRTRKRVYFRANVSLIPMTWKSSTYFIPKDSYDQAD
         VIVANPYLKGDPTLQYGRGEGKXITHEFPNLLTNNHITVSGRVEVYHVAHL
         RWGIDEVNVDQPFYISRNKTIETARCSITHIIGINVVRKCPGGSCITSLCRDSOTG
         LYEAKTCELPKRSOTAKESIMEMPSLHSTVECTEKTNTETAPNLQNMKNCKSTWDY
         IKNSDPQNTSPMTEMNPTHTPFTSLKSKQVQVLVDKSGMSAEDIYLLALIKI
         FKLIGNIT"
BASE COUNT 440 a 231 c 231 g 363 t
ORIGIN
Query Match      13.5%; Score 381; DB 3; Length 1265;
Best Local Similarity 64.3%; Pred. No. 8.5e-91;
Matches 640; Conservative 0; Mismatches 340; Indels 15; Gaps 4;
QY 5 TCACAGGAGAGGTACAGCAATGGGGCAATTAAGAAGTCTGTGTCATCTTGATCTTC 64
    |||
DB 89 TTACTGTAACAGTGTGCAAAATAGTGCTCTGCTGATATTTATCTGTCTTAACCTTGC 148
QY 65 ACCTTCTAAGAGGGCCCTGAGTAAATTCACCTATTCAGCTGAACAACAATGGCTATGAAG 124
```

Db 149 ATCTTCCCTGG--AAATGAAAGTTCAATGTAATTTGATTAACAAATGGGTATATG 205  
 QY 125 GCATTTGCTGTCATCGACCCCAATGTGCGAAGATGAAACCTCATCAACAAATA 184  
 Db 206 GCATTTGCTGTCATCGACCCCAATGTGCGAAGATGAAACCTCATCAACAAATA 265  
 QY 185 AGACATGGTGGCCAGGACCTCTGTATCTGTGTAAGTACAGAGAAAGGATTTAT 244  
 Db 266 AGGAATGGTAACTGAGCTTACTACTACTGTTTCAAGCCCAAGAGAGTTAT 325  
 QY 245 TCAAAATGTGTCATTTGATTCCTGAAAACATGGAAGCAAAAGCTGATGAGAC 304  
 Db 326 TCAAGATGTGACATTTTAAATGACCTGGAATGCAAAATCTAGACTCTATAC 385  
 QY 305 CAAACTGAGACTTCAAAATGCTGATGTTCTGGTTGCTGAGTCTACTCTCCAGTA 364  
 Db 386 CAAAGAGATCATATGACAGGACAGATGCTACTGTTTCAATCCCTATCAAAATATG 445  
 QY 365 ATGATGACCTTACACTGAGCAGATGGCAACTGTGAGAGAAAGGATGAAAGTCCAC 424  
 Db 446 GAGATGATCTTACACTGATGAAAGTGTGAGAGAAAGGATGAAAGTCCAC 505  
 QY 425 TCACTCTGATTTGATGCGAGAAAGTACTGATATGAGCACAAGGTAGGCGAT 484  
 Db 506 TTAATCCAACTTCTGTTGATGATTAATTTCCACATCTATGGTCCGAGGAGATAT 565  
 QY 485 TTGTCATGATGGGCTCATCTAGATGGGAGATTTGACAGTCAATATGATGAA 544  
 Db 566 TTGTCATGATGGGCTCATCTAGATGGGAGATTTGATGATGATGATGATGAGAC 625  
 QY 545 AATCTCTATCTC--AAATGAAAGTACAGAGATGATGATGATGATGATGATGATG 601  
 Db 626 CATCTATATTTCCAGAAAGAACTATGAGAGACAAAGTGTTCACATCATATCTAG 685  
 QY 602 GTACAAATGTAG--TAAAGAGTGCAGAGGAGGACGCTGTAACCAAAAGTGCACAT 658  
 Db 686 GTATTAATGTGTTTCAAGAAATGCTGAGGAGGACGCTGATTAACAGTCTATGAC 745  
 QY 659 TCAATTAAGTACAGAGCTTATGAAAGATGATGATGATGATGATGATGATGATG 718  
 Db 746 GTGACTACAGAGGCTGTATGAAAGAAATGATGATGATGATGATGATGATGATG 805  
 QY 719 CGGAAGGCTCTTATGATGTTGCAACAATGTTGATGATGATGATGATGATGATG 778  
 Db 806 CTGCAAGGATCTATGATGTTTATGCAAGTCTGATGATGATGATGATGATGATG 865  
 QY 779 AACAAACCAACAAGAGCTCCAAACAAAGATGATGATGATGATGATGATGATGATG 838  
 Db 866 AAAAAACACATACAGAGGCTCCAAACCAAGATGATGATGATGATGATGATGATG 925  
 QY 839 CATGGAATGATCCGTGATTTCTGAGACTTTAAGAAAACTCTATGACA-----A 892  
 Db 926 CATGGATGTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 985  
 QY 893 CACAGCCACCAATCCACCTTCTCATGCTGCAAGTGGCAAAAGATGATGATGATG 952  
 Db 986 ATCCACGACTATCTCATATTTCTATGCTCAAGTCCAAACAGGCGGATGATGATG 1045  
 QY 953 TCCTGCAATCTGGAAGATGCGACGTGATAC 987  
 Db 1046 TACTGTATTAATCTGGAAGATGCTGCAAGAGC 1080

RESULT 15  
 166494/c 166494 7218 bp DNA PAT 23-DEC-1997  
 LOCUS Sequence 14 from patent US 5670367.  
 DEFINITION  
 ACCESION 166494  
 NID 92724471  
 VERSION 166494.1 GI:2724471  
 KEYWORDS  
 SOURCE unknown.

ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 7218)  
 AUTHORS Dotner, F., Schefflinger, F. and Falkner, F. Gunter.  
 TITLE Recombinant fowlpox virus  
 JOURNAL Patent: US 5670367-A 14-23-SEP-1997;  
 FEATURES Location/Qualifiers  
 source 1..7218  
 BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others  
 ORIGIN  
 Query Match 1.8%; Score 52; DB 5; Length 7218;  
 Best Local Similarity 4.9%; Pred. No. 0.0023;  
 Matches 19; Conservative 211; Mismatches 156; Indels 0; Gaps 0;  
 QY 442 GCAGAAAAAGTTAGCTCAATATGAGCAGCAAGTAGGCAATTTGTCATGAGTGGCT 501  
 Db 1463 GTAGTAAAGATAGAGAAATTTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1404  
 QY 502 CATCTACATGGGAGTATTTGACAGATCAATATGATGAGAAATTTCTACTATCCAT 561  
 Db 1403 RRR 1344  
 QY 562 GGAAGATACAGCAGTAAGATGTCAGCAGATGATGATGATGATGATGATGATGATG 621  
 Db 1343 RRR 1284  
 QY 622 TGTACAGGAGGAGCTGTTTACCAAAAGATGCATTTCAATTAAGTAAACAGACTCAT 681  
 Db 1283 RRR 1224  
 QY 682 GAAAAAGATGATGTTGTTTCCAAATCCCGCAGAGGAGGCTCTATATATGTT 741  
 Db 1223 RRR 1164  
 QY 742 GCACAAATGTTGATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 801  
 Db 1163 RRR 1104  
 QY 802 CCAACAAAGCAAAATCAAAATGCA 827  
 Db 1103 RRR 1078

Search completed: August 5, 1999, 23:19:32  
 Job time: 8055 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:05:34 ; Search time 921.96 Seconds

(without alignments)  
763.363 Million cell updates/sec

Title: US-09-049-696-18

Perfect score: 2813  
Sequence: 1 GAAATCAGGAGGAGATGTCAC.....AATAATATCATTCCTTA 2813

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	780.8	27.8	878	1	T45884 Human colon specific
2	780.8	27.8	878	1	V16672 Polynucleotide seq
3	232	8.2	255	1	T22483 Human gene, signatu
4	70	2.5	398	1	X40511 Human secreted pro
5	40.2	1.4	5156	1	V10238 Arabidopsis ferula
6	38.6	1.4	5156	1	T90505 Arabidopsis thalia
7	37.8	1.3	110000	1	V21209_07 Continuation (8 of
8	36.8	1.3	7320	1	T68081 H. pylori cytoplas
9	35	1.2	1794	1	T33873 Candida RAP1 CDNA
10	34.8	1.2	6924	1	O67283 Human RPTP-beta. H
11	34.8	1.2	110000	1	X20248_00 Borrelia burgdorfe
12	34.4	1.2	110000	1	X20248_08 Continuation (9 of
13	34.2	1.2	2736	1	O84047 Insertion sequence
14	34	1.2	5319	1	V35235 R. prowazekii S-1a
15	34	1.2	110000	1	X20248_05 Continuation (6 of
16	33.8	1.2	3447	1	T08711 Sak-a serine-threo
17	33.6	1.2	1200	1	O52267 Sequence encoding
18	33.6	1.2	5391	1	V71729 Upstream sequence
19	33.6	1.2	5407	1	V71738 Upstream sequence
20	33.6	1.2	286	1	X39448 Human secreted pro
21	33.4	1.2	2312	1	V21646 T. gondii chorioma
22	33.2	1.2	2520	1	Q77884 Neutral thread prot
23	33.2	1.2	2520	1	T27766 AD 16c human neuro
24	33.2	1.2	825	1	T83807 DNA encoding a sur
25	33	1.2	2980	1	N91071 Nucleotide sequenc
26	33	1.2	2615	1	N81002 Bacillus thuringie
27	33	1.2	2980	1	N82139 Cryc gene. New gen
28	33	1.2	2983	1	Q71026 CryIIa gene which
29	33	1.2	1984	1	T13030 Cotton fibre-speci
30	33	1.2	2615	1	T33187 Bacillus thuringie
31	33	1.2	1985	1	T30250 Cotton fibre clone
32	33	1.2	1985	1	T70036 Cotton H6 gene and
33	33	1.2	678	1	V86928 EST clone BK296. N
34	32.8	1.2	1534	1	T15995 Human interleukin-
35	32.8	1.2	1838	1	T90504 Arabidopsis thalia
36	32.8	1.2	1838	1	V10237 Arabidopsis ferula
37	32.8	1.2	3342	1	V52339 Streptococcus pneu
38	32.8	1.2	110000	1	V21209_13 Continuation (14 o
39	32.6	1.2	698	1	V06140 Viral infection ge
40	32.6	1.2	799	1	X13748 Enterococcus faeca
41	32.4	1.2	1423	1	V19156 Human XMG growth f
42	32.4	1.2	1746	1	V68522 Nucleotide sequenc
43	32.4	1.2	15363	1	X13216 Enterococcus faeca

## ALIGNMENTS

44 32.4 1.2 4137 1 X20278  
45 32.2 1.1 10670 1 T99449

Borrelia burgdorfe  
Staphylococcus car

## RESULT 1

ID T45884 T45884 standard; cDNA; 878 BP.

AC T45884;

DT 13-MAR-1997 (first entry)

DE Human colon specific gene CS65 cDNA partial clone.

KW Colon specific gene; CS65; colon cancer; metastasis; diagnosis;

OS Homo sapiens.

FT cds

FT 2..691

FT /\*tag= a

PN MO9639419-A1.

PD 12-DEC-1996.

PF 06-JUN-1995; 007289.

PA 06-JUN-1995; MO-T07289.

PI (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Yu G;

PI WPI; 97-043054/04.

PI P-PSDB; W06548.

PI Human colon specific genes and their expression products - detection

PI of which, in non-colon tissue samples, can be used as indication of

PI colon cancer metastasis

PS claim 1: Fig 5: 60pp: English.

CC 13 cDNA clones (T45880-92), most of them partial clones, correspond

CC to human colon specific genes, designated CS61, CS62, etc., that

CC are primarily expressed in tissues derived from the colon. CS67

CC and CS610 show reduced expression in colon cancer cells as compared

CC to that in normal cells; the remaining genes are overexpressed in

CC colon cancer. The partial cDNA sequences can be used to isolate

CC full-length clones and genomic clones including the complete gene.

CC CS6 nucleic acids can be used to produce CS6 polypeptides (see also

CC W06545-53) in transformed host cells, as probes to detect disorders

CC of the colon, partic. colon cancer and colon cancer metastasis, and

CC in gene therapy.

SO Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 27.8%; Score 780.8; DB 1; Length 878;

Best local Similarity 98.9%; Pred. No. 6.9e-221;

Matches 796; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 1992 TGTCTACTCAAGGATATTTCAACTATGACACGAATGATGATACAGTGAAGTCCG 2051

DB 1 TGTCTACTCAAGGATATTTCAACTATGACACGAATGATGATACAGTGAAGTCCG 60

QY 2052 GGCTCGGGAGGAGTTAAGCAGCAGAGGAGAGATGATACCCAGAGAGGAGGACT 2111

DB 61 GGCTCGGGAGGAGTTAAGCAGCAGAGGAGAGATGATACCCAGAGAGGAGGACT 120

QY 2112 GTACATACCTGCTGATGATGAGATGATGATGATGATGATGATGATGATGATGAT 2171

DB 121 GTACATACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180

QY 2172 TAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2231

DB 181 TAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240

QY 2232 ATTTGGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2291

DB 241 ATTTGGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

QY 2292 CACCGACTGTAAGGCGGAATTCACGGGGGAGTCTCATTAATGACTTGACAGCTCC 2351

DB 301 CACCGACTGTAAGGCGGAATTCACGGGGGAGTCTCATTAATGACTTGACAGCTCC 360

OY	2352	TGGGGATGATTATGACCATGGAACAGCTCACAAGTATTCATTGTGAATGAAGTACAAAGTAT	2411
Db	361	TGGGGATGATTATGACCATGGAACAGCTCACAAGTATTCATTGTGAATGAAGTACAAAGTAT	420
OY	2412	TCTTGATCTCGAGACAGAAGTTCAATGATCTCTTCGAAGTAATACTACTGCTCTCATCC	2471
Db	421	TCTTGATCTCGAGACAGAAGTTCAATGATCTCTTCGAAGTAATACTACTGCTCTCATCC	480
OY	2472	AAAGGAAGCCAACTCTGAGGAAGTCTTTTTGTTTAAACCAGAAAACATTACTTTGGAAA	2531
Db	481	AAAGGAAGCCAACTCTGAGGAAGTCTTTTTGTTTAAACCAGAAAACATTACTTTGGAAA	540
OY	2532	TGGCACAGATCTTTTCATTCTCTAATCAGGCTGTTTAAGTGCATCTGGAATCGAAT	2591
Db	541	TGGCACAGATCTTTTCATTCTCTAATCAGGCTGTTTAAGTGCATCTGGAATCGAAT	600
OY	2592	ATCCAACATTCGACGAGATCTTGTATTATCTCCACAGACTCCGCCAGAGACACTAG	2651
Db	601	ATCCAACATTCGACGAGATCTTGTATTATCTCCACAGACTCCGCCAGAGACACTAG	660
OY	2652	TCCCTGATGAAAAGCTCTGCTCTTGT - CCTAATATTCATATCAAAGCNCATTCCTGGCA	2710
Db	661	TCCCTGATGAAAAGCTCTGCTCTTGT - CCTAATATTCATATCAAAGCNCATTCCTGGCA	720
OY	2711	TTTCACATTTTAAAAATTATGTGGAAGTGAATAGAGAGAACTGACGTCTGCATATAGCTTAGG	2770
Db	721	TTTCACATTTTAAAAATTATGTGGAAGTGAATAGAGAGAACTGACGTCTGCATATAGCTTAGG	780
OY	2771	GCTGATTTTGTGTCGATMAATAAA 2795	
Db	781	GCTGATTTTGTGTCGATMAATAAA 805	
RESULT	2		
ID	VI6672	VI6672 standard; cDNA; 878 BP.	
AC	VI6672:		
DT	22-JUN-1998	(first entry)	
DE	Poly nucleotide sequence of a colon-specific gene.		
KM	Colon-specific gene; probe; detection; expression; human;		
.KW	diagnostic assay; colon cancer; antibody; screening; ss.		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	CDS	2..685	
FT		/tag= a	
FT		/note= "no stop codon given"	
PN	US5733748-A.		
PD	31-MAR-1998.		
PF	06-JUN-1995; 469667.		
PR	06-JUN-1995; US469667.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Rosen C, Yu G;		
P1	WPI: 98-229823/20.		
DR	P-P.SDB: M46879.		
PT	cancer specific nucleic acids - useful as probes for detecting colon		
PT	cancer micrometastases		
PS	Claim 15; Fig 5A-B; Snp; English.		
CS	VI668-81 represent polynucleotide sequences of partial or full-length		
CC	cDNA clones of colon-specific genes. The polynucleotides can be used		
CC	as probes to detect expression of the corresponding human genes, e.g., in		
CC	diagnostic assays for detecting micrometastases of colon cancer.		
CC	Recombinant cells containing the polynucleotides can be used to		
CC	produce the polypeptides, in order that antibodies can be raised and		
CC	used in further screening or diagnostics.		
Sequence	878 BP; 257 A; 179 C; 188 G; 241 T;		

	Query Match	27.8%;	Score 780.8;	DB 1;	Length 878;
	Best Local Similarity	98.9%;	Pred. No. 6.9e-221;		
	Matches 796;	Conservative	0;	Mismatches 8;	Indels 1;
					Gaps 1;
0Y	1992 TGTCTACTCAAGGTATTTCACAACTTATACACAGATGATGATATACAGTGTAAAGTCGC	2051			

Db	1	TGTTACTCAAGGATATTTCACAACTATAGACACCAATAGTAGATACAGATGTAAGTAAAGTGGC	60
Oy	2052	GGCTCTGGGAGGAGTTTAACGCACGCCAGAGAGATATACCCAGCAGAGTGGACCACT	2111
Db	61	GGCTCTGGGAGGAGTTTAACGCACGCCAGAGAGATATACCCAGCAGAGTGGACCACT	120
Oy	2112	GTACATACCTGGCTGGATTTGAAATGATGAATAATCAATGGAATCCACCAAGACCTGAAT	2171
Db	121	GTACATACCTGGCTGGATTTGAAATGATGAATAATCAATGGAATCCACCAAGACCTGAAT	180
Oy	2172	TAAATAAGATGATGTTCACACAGACCAAGTGTGTTTCAGCAGAAATCTCTGGAGGCTC	2231
Db	181	TAAATAAGATGATGTTCACACAGACCAAGTGTGTTTCAGCAGAAATCTCTGGAGGCTC	240
Oy	2232	ATTGTGGCTTCTGATGTGCCAAATGTCTCCATACCTGATCTCTTCCAGCTGGCCAAAT	2291
Db	241	ATTGTGGCTTCTGATGTGCCAAATGTCTCCATACCTGATCTCTTCCAGCTGGCCAAAT	300
Oy	2292	CACCGACCTTAAGGCGGAAATTCACGGGGGAGCTCATTAATCTGCATTTGGACAGCTCC	2351
Db	301	CACCGACCTTAAGGCGGAAATTCACGGGGGAGCTCATTAATCTGCATTTGGACAGCTCC	360
Oy	2352	TGGGATGATTTATGACCATGGAACAGCTCACAAGTATATCATTTGGAATTAAGTAAAGTAT	2411
Db	361	TGGGATGATTTATGACCATGGAACAGCTCACAAGTATATCATTTGGAATTAAGTAAAGTAT	420
Oy	2412	TCTTGTATCTGAGAGACAAAGTTCAAATACCTGTAAGGAATTAAGTAAAGTAAAGTAT	2471
Db	421	TCTTGTATCTGAGAGACAAAGTTCAAATACCTGTAAGGAATTAAGTAAAGTAAAGTAT	480
Oy	2472	AAGGAGACCAACTCTAGAGAAAGTCTTTTGTGTTAAACCGAAGAAACATTAAGTAAAGT	2531
Db	481	AAGGAGACCAACTCTAGAGAAAGTCTTTTGTGTTAAACCGAAGAAACATTAAGTAAAGT	540
Oy	2532	TGGCAGACATCTTTTCATTTGCTATATTCAGGCTGTGTAAGTGCATTTGAATTCAGAAAT	2591
Db	541	TGGCAGACATCTTTTCATTTGCTATATTCAGGCTGTGTAAGTGCATTTGAATTCAGAAAT	600
Oy	2592	ATTCACAACATGAGAGAGATCTTTGTTATCTCTCCACAGACTTCGGCCAGACACACTAG	2651
Db	601	ATTCACAACATGAGAGAGATCTTTGTTATCTCTCCACAGACTTCGGCCAGACACACTAG	660
Oy	2652	TCTGTATGAAAGCTGTGCTCTGT-CTATAATTCATATCAACAGACACCAATCTCTGGCA	2710
Db	661	TCTGTATGAAAGCTGTGCTCTGTGCTATATTCATATCAACAGACACCAATCTCTGGCA	720
Oy	2711	TTTCACATTTAAAAATTTATGTGGAAGTGGGTAGGAGAACTGCAGTTGTCAATAGNCTAGG	2770
Db	721	TTTCACATTTAAAAATTTATGTGGAAGTGGGTAGGAGAACTGCAGTTGTCAATAGNCTAGG	780
Oy	2771	GCTGAATTTTGTCAAGATTAATAA 2795	
Db	781	GCTGAATTTTGTGCGGAGTAATAA 805	

RESULT	3
ID	T22483
NC	T22483 standard; cDNA to mRNA; 255 BP.
DT	22-AUG-1996 (first entry)
DE	Human gene signature HUMS04094.
KW	gene signature; messenger RNA; mRNA; relative abundance; frequency
KW	human; cloning; mapping; non-biased library; diagnosis; detection;
KW	cell typing; abnormal cell function; ss.
OS	Homo sapiens.
PN	MO9514772-A1.
PD	01-JUN-1995.
PF	11-NOV-1994; J01916.
PR	12-NOV-1993; JP-355504.
PA	(MATS/) MATSUBARA K.
PA	(OKUB/) OKUBO K.
PI	Matsubara K; Okubo K;
DR	WPI; 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 PS Claim 1: Page 1138-1139; 2245pp; Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in 119001-176837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNA species. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SQ Sequence 235 BP; 76 A; 52 C; 39 G; 80 T;

Query Match  
 Best Local Similarity 97.5%; Pred. No. 6,1e-59;  
 Matches 232; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2575 GATCTGAAATCAGAAATATCCAACTGACGAGTATCTTTGTTATCTCCACAGACT 2634  
 DB 1 GATCTGAAATNAGAAATATCCAACTGACGAGTATCTTTTATNCTCCACAGACT 60

QY 2635 CCGCCGAGACACACTAGTCTCTGATGAAGCTGCTCTCTCTCTATATATATCAAC 2694  
 DB 61 CCGCCGAGACACACTAGTCTCTGATGAAGCTGCTCTCTCTCTATATATCAAC 120

QY 2695 AGCACCATTCTGTCGATTCACATTTTAAATATGATGAGATGAGAGACTGAC 2754  
 DB 121 AGCACCATTCTGTCGATTCACATTTTAAATATGATGAGATGAGAGACTGAC 180

QY 2755 CTTGTCATATAGCTAGGCTGATTTTGTGATTAATAATAATCATTCATCTT 2812  
 DB 181 CTTGTCATATAGCTAGGCTGATTTTGTGATTAATAATAATCATTCATCTT 238

RESULT 4  
 X40511  
 ID X40511 standard; cDNA; 398 BP.  
 AC X40511:  
 DT 18-JUN-1999 (first entry)  
 DE Human secreted protein 5' EST SEQ ID No: 111.  
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
 OS Homo sapiens.  
 PN WO9906550-A2.  
 PD 11-FEB-1999.  
 PE 31-JUL-1998; IB1232.  
 PR 01-AUG-1997; US-905144.  
 PA (GEST) GENSET.  
 PI Duclet A, Dumas Milne Edwards J, Lacroix B;  
 DR WPI: 99-153780/13.  
 DR P-PSDB: Y11789.  
 PT New isolated prostate-derived nucleic acids - used to develop  
 PT products which may have cytokine, immune regulatory, haematopoiesis  
 PT regulating, anti-inflammatory or tumour inhibition activity  
 PS Claim 1: Page 228; 675pp; English.  
 CC X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human  
 CC secreted proteins expressed in prostate, and encode the proteins given in  
 CC Y1116 to Y1193 respectively. The proteins given represent the signal  
 CC peptide and an N-terminal fragment of a secreted protein. The nucleic

CC acid sequences can be used for producing secreted human gene products.  
 CC They can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell proliferation and  
 CC differentiation activity, haematopoiesis regulating activity, tissue  
 CC growth regulating activity, reproductive hormone regulating activity,  
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
 CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition  
 CC activity or other activities. The products can be used in forensic, gene  
 CC therapy and chromosome mapping procedures. The sequences can also be used  
 CC for obtaining corresponding promoter sequences. The nucleic acids  
 CC encoding the signal peptide can be used for directing extracellular  
 CC secretion of a polypeptide or the insertion of a polypeptide into a  
 CC membrane, or importing a polypeptide into a cell.  
 SQ Sequence 398 BP; 123 A; 86 C; 76 G; 113 T;

Query Match  
 Best Local Similarity 62.6%; Pred. No. 6,2e-11;  
 Matches 109; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 98 TTCAGCTGAACAACATGAGCTATGAGGATGCTGTCATATGACCCCAATGTCAG 157  
 DB 220 TACAGCTGAACAACATGAGCTATGAGGATGCTGTCATATGACCCCAATGTCAG 279

QY 158 AGATGAACACATTCATCAAAATTAAGACATGTCAGCCAGCATCTGTATCTGT 217  
 DB 280 AGATGAACACATTCATCAAAATTAAGACATGTCAGCCAGCATCTGTATCTGT 339

QY 218 TGAAGCTACAGGAAGGATTTTATTCAAAATGTCATTTGATTCCTG 271  
 DB 340 TTAATCTACCAAGACAGATATTTTTCAGAAATTAAGATTTTAATACCTG 393

RESULT 5  
 VI0238/c  
 ID VI0238 standard; cDNA; 5156 BP.  
 AC VI0238:  
 DT 20-JUL-1998 (first entry)  
 DE Arabidopsis ferulate-5-hydroxylase genomic DNA.  
 KW ferulate-5-hydroxylase; cinamate-4-hydroxylase;  
 KW tissue-specific promoter; ligand; syringyl; delignification;  
 KW pulp; transgenic plant; ss.  
 OS Arabidopsis thaliana.  
 FH key  
 FT CDS Location/Qualifiers  
 FT 2487..4522  
 FT /\*tag= a  
 FT /\*note= "contains introns"  
 FT 2997..3125  
 FT /\*tag= b  
 FT /\*number= 1  
 FT 3557..3901  
 FT /\*tag= c  
 FT /\*number= 2  
 PN WO9803535-A1.  
 PD 29-JAN-1998.  
 PE 18-JUL-1997; U12624.  
 PR 16-DEC-1996; US-032908.  
 PR 19-JUL-1996; US-022228.  
 PA (PURD) PURDUE RES FOUND.  
 PI Chapple CS;  
 DR WPI: 98-120689/11.  
 DR P-PSDB: W40099.  
 PT Tissue-specific regulatory plant promoter linked to  
 PT ferulate-5-hydroxylase gene - useful for manipulation of plant  
 PT lignin composition  
 PS Claim 14: Page 48-51; 71pp; English.  
 CC This genomic DNA sequence includes a coding region, interrupted  
 CC by 2 introns, that codes for the ferulate-5-hydroxylase (F5h, see  
 CC W40099) of Arabidopsis thaliana. It comprises the HindIII-XhoI  
 CC fragment of pBIC20-F5h genomic clone isolated from a pBIC50  
 CC cosmid library using F5h cDNA (see VI0237) as probe. F5h is a  
 CC cytochrome P450-dependent monooxygenase that catalyses the  
 CC meta-hydroxylation of ferulic acid in the monomer-specific pathway



QY 403 GAGAGGGTGAAGATCCACCTCACTCCTGATTTATTCGACAGAAAAGTTAGCTGAA 462  
DB 66736 GAAAGGGCTGAAGTACTGCTGCTAAATGATGTTATCAAGAAATTTATTTGAT 66795  
QY 463 TATGACCCCAAGATGGGATTTGTCATGATGAGGCTCATGACATGGGAGTATTT 522  
DB 66796 ATTGACCCCAAGATGGGATTTGTCATGATGAGGCTCATGACATGGGAGTATTT 66855  
QY 523 GACGACTACATTAATGATGAGAAATTTCTACTTATCCATGGAAGATACAGCAGTAAGA 582  
DB 66856 GCTCCATATGACTATATGTTGATTAACGACCAATATCATTTAAGGAATGGAGAA 66915  
QY 583 TGTTCAGCAGGATTTACTGCTACCAATGTATGAAGA 619  
DB 66916 ATAAAGTTGTTAAATGGAAGACTTAAGATTAAGA 66952

RESULT 8  
ID T68081 standard; DNA; 7320 BP.  
AC T68081;  
DE 16-JUL-1997 (first entry)  
KW H. pylori cytoplasmic protein ORF 119610309orf1.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.  
OS Helicobacter pylori.  
FH Key Location/Qualifiers  
FT cds 1..7320  
FT /tag= a  
FT /note= "no stop codon given"

PN W09640893-A1.  
PD 19-DEC-1996.  
PE 06-JUN-1996; U09122.  
PR 07-JUN-1995; US-487032.  
PR 01-APR-1996; US-630405.  
PA (ASTR) ASTRA AB.  
PI Berghindh OT, Smith D, Mellgaard BL;  
DR P-PSDB; W20828.  
PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
PT Claim 9; Page 887-889; 1481pp; English.  
CC The present sequence encodes a Helicobacter pylori cytoplasmic protein  
CC that may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds.  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
SQ Sequence 7320 BP; 2828 A; 1317 C; 1436 G; 1739 T;

Query Match 1.3%; Score 36.8; DB 1; Length 7320;  
Best Local Similarity 46.8%; Pred. No. 2.1;  
Matches 116; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 775 ACAGAACAAACACACAAAGAGCTCCAAAGCAAAATTAATTAATTCATTCGCA 834  
DB 1288 ACCACACAAAGAACACACACAAAGAGCTCCAAAGCAAAATTAATTCATTCGCA 1347  
QY 835 AGCAGATGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 894  
DB 1348 ACCACACAAAGAACACACACAAAGAGCTCCAAAGCAAAATTAATTAATTCATTCGCA 1407

QY 895 GAGCACAATATCCACCTTCATTCGTCGACATTTGACAAAGATGTGTGTAGTC 954  
DB 1408 AAACAAAGAAACCAATTAAGCGACGACATCCCTTAATTCCTAAATTAATTTAAACCG 1467  
QY 955 CTGACAAATTCGACATGGGAGCTGGTGAACCGCTCAATGATGATCAAGCAGCG 1014  
DB 1468 AGCGAAGAAAGTTTAAATCAAGGAGCTAAACTCGCTACAGGCTTAACATTAAGCCATTG 1527  
QY 1015 CAGCTTTT 1022  
DB 1528 GAGCTTTT 1535

RESULT 9  
ID T33873 standard; cDNA; 1794 BP.  
AC T33873;  
DE 24-NOV-1996 (first entry)  
KW Candida RAPT1 cDNA.  
KW RAPT1; rapamycin binding protein; FKBP; immunosuppressive; agonist;  
KW fungicide; anti-mycotic; antagonist; cell proliferation; ss; ds.  
OS Candida albicans.  
FH Key Location/Qualifiers  
FT cds 1..1689  
FT /tag= a

PN W09533052-A1.  
PD 07-DEC-1995.  
PE 30-MAY-1995; U06722.  
PR 27-MAY-1994; US-250795.  
PR 20-DEC-1994; US-360144.  
PA (MITO-) MITOTIX INC.  
PI Berlin V, Chiu MT, Cottarel G, Damaguez V;  
DR WPI; 96-030563/03.  
DR P-PSDB; W04236.  
PT Rapamycin binding protein RAPT-1 - used as agonist, or antagonist of  
PT rapamycin cellular proliferation regulation  
PT Example 11; Page 92-95; 121pp; English.  
CC A cDNA clone (T33873) codes for a partial sequence (W04236) of a  
CC RAPT1-like protein of Candida albicans. It was isolated from C.  
CC albicans cDNA in lambda ZAP by PCR amplification using primers  
CC (T33887-88) based on conserved regions of RAPT1 and TOR proteins  
CC (see also W04234-35 and W04237-38) and use of the PCR product to  
CC probe C. albicans cDNA libraries. The cDNA can be used for the  
CC prodn. of recombinant RAPT1-like protein, esp. its rapamycin  
CC binding domain, or to design probes/primers used to detect  
CC Candida infection.  
SQ Sequence 1794 BP; 585 A; 273 C; 418 G; 518 T;

Query Match 1.2%; Score 35; DB 1; Length 1794;  
Best Local Similarity 54.2%; Pred. No. 3.2;  
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 691 TGTGATTTGTTCTCCATTCCTCCGACGAGAGAGGCTTCTATATGTTGCACACAT 750  
DB 1372 TGGGGTTGTTGATTTCCCAAGAGGCTTGATCAACGGATATAGCTGTTCCAAA 1431  
QY 751 GTTGATTTATGTTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 810  
DB 1432 GTCAACACTGCGAATATTAACGACAGACAGATTTGACGAAAAAGCAACTGTAAATTTG 1491  
QY 811 CAAATCAAAA 821  
DB 1492 CAAAGCAAAA 1502

RESULT 10  
ID 067283 standard; cDNA; 6924 BP.  
AC 067283;  
DE 29-MAR-1995 (first entry)  
DE Human RPTP-beta.

RPRP-beta: receptor protein tyrosine phosphatase-beta; mapping: gene therapy; ss.  
 OS Homo sapiens.  
 FT cds Location/Qualifiers  
 FT 1..6924  
 FT /tag= a  
 FT signal\_peptide 1..72  
 FT /tag= b  
 FT mat\_peptide 73..6924  
 FT /tag= c  
 PN M09418308-A.  
 PD 18-AUG-1994.  
 PF 09-FEB-1994; U01541.  
 PR 10-FEB-1993; US-015973.  
 PA (UNIV) UNIV NEW YORK STATE.  
 PI Schlessinger J;  
 DR WPI: 94-279729/34.  
 DR P-PDB: R57902.  
 PT Human receptor-type protein tyrosine phosphatase-beta  
 PT protein/glycoprotein - useful in screening assay for compounds,  
 PT e.g. drugs, inhibiting or stimulating the protein  
 PS Disclosure: Fig.1A-1L; 101pp; English.  
 CC CDNA clones containing the coding sequences of RPRP-beta were  
 CC isolated from a lambda-gt11 human infant brain stem library. The  
 CC corresponding gene was mapped to human chromosome 7q31-q33.  
 SO Sequence 6924 BP; 2063 A; 1482 C; 1440 G; 1939 T;

Query Match 1.2%; Score 34.8; DB 1; Length 6924;  
 Best Local Similarity 48.5%; Pred. No. 8;  
 Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

OY 53 TCTTGATTTTCACCTCTGAGAGGGCCCTGAGTATTCATCTATTCATCTGAGACACA 112  
 DB 5927 TCTTCAATTCATGATACACTGCTGAGCCATCTAGTAAAGAACTGAGGCTGACACA 5986  
 OY 113 ATGCGATGAGGAGCTGCTGTCGATGCAATGCAATGCGCCATGCGCAAGATGAACACTCA 172  
 DB 5987 GTCATATTCATGCTGCTATGATGCACTCTCTGATGCGCCAGGACGCAAAACAAAGC 6046  
 OY 173 TTCAACAATTAAGAGACATGTCAGCCAGCATCTCTGATCTGTTGAAGTACAGGAA 232  
 DB 6047 TAGGAAGAACATTCACGCTCTGAGCGATCAATATACAGCAGAGTATCTGCGAG 6106  
 OY 233 AGCGATTTATTCAAA 250  
 DB 6107 CCCTAAGCAATGCAACA 6124

RESULT 11  
 X20248\_00  
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
 WP Fragment Name Begin End  
 WP X20248\_00 1 110000  
 WP X20248\_01 100001 210000  
 WP X20248\_02 200001 310000  
 WP X20248\_03 300001 410000  
 WP X20248\_04 400001 510000  
 WP X20248\_05 500001 610000  
 WP X20248\_06 600001 710000  
 WP X20248\_07 700001 810000  
 WP X20248\_08 800001 910000  
 WP X20248\_09 900001 910715  
 ID X20248 standard; DNA: 910715 BP.  
 AC X20248;  
 DT 04-MAY-1999 (first entry)  
 DE Borrelia burgdorferi polynucleotide sequence #1.  
 KW Borrelia burgdorferi; Spirochaete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.  
 OS Borrelia burgdorferi.  
 PN M09858943-A1.  
 PD 30-DEC-1998.

PF 18-JUN-1998; U12764.  
 PR 03-SEP-1997; US-057483.  
 PR 20-JUN-1997; US-050359.  
 PR 22-JUL-1997; US-053344.  
 PR 22-JUL-1997; US-053377.  
 PA (HUMAN) HUMAN GENOME SCI INC.  
 PA (MED-) MEDIMONE INC.  
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,  
 PI White OR;  
 DR WPI: 99-081217/07.  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease  
 PS Claim 1: Page 157-671; 1128pp; English.  
 CC X20248 to X20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.  
 SO Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T;

Query Match 1.2%; Score 34.8; DB 1; Length 110000;  
 Best Local Similarity 46.6%; Pred. No. 38;  
 Matches 111; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

OY 2498 TTTTGTTTAAACGAGAAACATTTCTTTGAAATGCGACAGATCTTTCTATTCCTATTC 2557  
 DB 13611 TTTTCATTAACCTAGTGAATTTCTCAAAAATATCAATAAACAAATTTCTGATT 13670  
 OY 2558 AGCGTGTGATAGGTGATCTGAATCAAGAAATATCCACATTCGACGATCTTGT 2617  
 DB 13671 CAATGGTTTATTTTATTTATTTGTTGACCAATTAATAACCATGATTATGAACATGAT 13730  
 OY 2618 TTATTCCTCCAGACAGCTCCGCGAGACACCTAGTCTGATGAACGCTGCTCTGTC 2677  
 DB 13731 TTAAGTTCTTGAATCTCCCTTAATATCATCATATAGATGTTAAAGCTACAAATG 13790  
 OY 2678 CTATATTCATATCAACAGCACCATTCCTGCGATTCATTTTAAATATATGAGAA 2735  
 DB 13791 AATATATTATATCAAAATTAACAAATAGCTTAGACGTATGAATAAATTAATTGGCA 13848

RESULT 12  
 X20248\_08  
 Continuation (9 of 10) of X20248 from base 800001 (Borrelia burgdorferi polynucleotide  
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
 WP Fragment Name Begin End  
 WP X20248\_00 1 110000  
 WP X20248\_01 100001 210000  
 WP X20248\_02 200001 310000  
 WP X20248\_03 300001 410000  
 WP X20248\_04 400001 510000  
 WP X20248\_05 500001 610000  
 WP X20248\_06 600001 710000  
 WP X20248\_07 700001 810000  
 WP X20248\_08 800001 910000  
 WP X20248\_09 900001 910715  
 Query Match 1.2%; Score 34.4; DB 1; Length 110000;  
 Best Local Similarity 57.4%; Pred. No. 50;  
 Matches 62; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
 OY 2706 TGGCATTCACATTTTAAATATATGATGAGATAGAGACTGCGAGCTGCTCAATAGC 2765  
 DB 22712 TAGGTGACAAATATATATCTTATCTGTTTCAATAGGCGTTTCAAAATTCATTTGCG 22771  
 OY 2766 CTAGGCTGAATTTTGTGACATTAATAATAATATCATTCCTTA 2813



Db 22772 GCATGTTTGGATTTTTCAGACAGTAAGAAAAGCCCATTTTCA 22819

## RESULT 13

084047/c  
ID 084047 standard; DNA: 2736 BP.  
AC 084047;  
DE 24-OCT-1995 (first entry)  
DT Insertion sequence pHG 35-5-17 comprising GCAP-1 gene.  
KW Guanylate cyclase activating peptide; GCAP-1; cyclic GMP; disease;  
KW treatment: kidney; bowel; gastrointestinal; respiratory; urogenital;  
KW circulatory; nervous system; sensory organs; integument; deficiency;  
KW bone; osteoporosis; teeth; ds.  
OS Homo sapiens.  
PN WO9503410-A.  
PD 02-FEB-1995.  
PF 20-JUL-1994: E02391.  
PR 20-JUL-1993: DE-324290.  
PA (FORS/) FORSMANN W.  
PI Adermann K, Hill O, Kuhn M, Maegert H, Pardigol A;  
PI Ralida M, Schulz-Knappe P, Zucht H;  
DR WPI: 95-075237/10.  
PT New gene for guanylate cyclase activating peptide - and its  
PT expression products, useful in diagnosis, treatment and gene  
PT therapy of eg kidney and bowel disease, osteoporosis etc.  
PS Claim 4: Page 34-37: 76pp; German.  
CC Guanylate cyclase activating peptide (GCAP-1) is an activator of  
CC cyclic GMP production in cells. The peptide and active fragments of  
CC it are useful for treating diseases associated with disturbed  
CC electrolyte water balance, of the gastrointestinal, respiratory,  
CC urogenital, circulatory and nervous systems; involving the  
CC integument and sensory organs; and for substitution therapy  
CC of diseases involving a deficiency of GCAP-1 and diseases involving  
CC bone construction (osteoporosis) and teeth. They are used  
CC especially during intensive care of acute disease.  
SQ Sequence 2736 BP: 725 A; 527 C; 602 G; 842 T;

Query Match 1.2%; Score 34.2; DB 1; Length 2736;  
Best Local Similarity 51.7%; Pred. No. 7.1;  
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 2486 CTGAGCAAGCTCTTTTGTAAACGAAAACCTTTTGAATAATGCACAGATCTT 2545  
DB 2621 CTGAGGACAGAGATGCTTGAACCGGAGGAGGTTTGCAGCGGATCGTG 2562  
QY 2546 TCATTCCTATTCAGGCTGTGATAGTGCATCTGAATCAGAAATATCCACATTCGAC 2605  
DB 2561 CCACCTCCTCCAGCTGGGCAACAGGCAAGCTCATCTCAAAAAGAAATATNA 2502  
QY 2606 GAGTATCTTTGTTTATCTCCACA 2630  
DB 2501 TAATAGTCTGTGCTTTCAGACACA 2477

## RESULT 14

V35235  
ID V35235 standard; DNA: 5319 BP.  
AC V35235;  
DE 28-SEP-1998 (first entry)  
DT R. prowazekii S-layer protein genomic DNA.  
KW Surface layer protein; S-layer; vaccine; typhus; spotted fever;  
KW infection; diagnosis; disease; ds.  
OS Rickettsia prowazekii.  
FH Key location/Qualifiers  
FT -35\_signal 340..345  
FT /tag= a  
FT -10\_signal 363..368  
FT /tag= b  
FT RBS 379..386  
FT /tag= c  
FT CDS 391..5229

FT /tag= d  
FT /product= S-layer protein  
FT stem\_loop 5270..5306  
FT /tag= e

PN US5783441-A.  
PD 21-JUL-1998.  
PF 20-DEC-1993: 169927.  
PR 09-AUG-1993: US-169927.  
PA (USNA) US SEC OF NAVY.  
PI Carl M, Ching W, Dasch GA, Dobson ME;  
DR WPI: 98-427031/36.  
PT Recombinant DNA encoding Rickettsia surface layer proteins - useful  
PT for diagnosing typhus and spotted fever and for preparing vaccines  
PT against them  
PS Claim 1: Column 11-24: 20pp; English.  
CC This sequence encodes the surface layer (S-layer) protein from  
CC R. prowazekii strain Breinl. This sequence is useful for vaccination  
CC against typhus and spotted fever rickettsial infection or for diagnosing  
CC diseases caused by these bacteria. The surface layer protein antigens can  
CC be produced recombinantly in large quantities.  
SQ Sequence 5319 BP: 1815 A; 766 C; 983 G; 1755 T;

Query Match 1.2%; Score 34; DB 1; Length 5319;  
Best Local Similarity 51.3%; Pred. No. 12;  
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 325 AATGCGATGTTCTGCTGCTGCTACTCCAGTAATGAGAACCCCTACACTGAG 384  
DB 1810 AATATATATGCTCTTCCACTGCTTATTCAGTTAGAGGAGTCTTAATACCGGT 1869  
QY 385 CAGATGGCACTGTGAGAGAGGAGGTAAGGATCCACTCCTCATTTTCATTCGA 444  
DB 1870 GATATAGTAAAGCTGTGTTATGCTGTACCAACATTAATTACTTAACGATGT 1929  
QY 445 GGAATAAGTTAGCTGAATATGACCAAGGTA 478  
DB 1930 TCATAATATATAGCACTGATGCGCAATATTA 1963

## RESULT 15

X20248\_05  
Continuation (6 of 10) of X20248 from base 500001 (Borrelia burgdorferi polynucleotide  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 210000  
WP X20248\_02 310000  
WP X20248\_03 410000  
WP X20248\_04 510000  
WP X20248\_05 610000  
WP X20248\_06 710000  
WP X20248\_07 810000  
WP X20248\_08 910000  
WP X20248\_09 910001  
WP X20248\_10 910715

Query Match 1.2%; Score 34; DB 1; Length 110000;  
Best Local Similarity 53.4%; Pred. No. 65;  
Matches 95; Conservative 0; Mismatches 80; Indels 3; Gaps 1;

QY 2445 TCAAGTAATGCTCTCATCCCAAGAGCAACTGTGAGAGTCTTTTGT 2504  
DB 80460 TCAAGCAATGCTTATGATGCTTTTCAAGATGATCAAAATCAGGAAATGAATT 80519  
QY 2505 TAAACCAAAAACATTACTTTGAAAATGCGACAGAT---CTTTTCATTCCTATTCAGGC 2561  
DB 80520 TAATGAATTAAGATATTTCACAAATTTTCAGTAGGTTGTTATGCTTAATATGG 80579  
QY 2562 TGTGATAGTGCATCTGAATTCAGAAATATTCACATTCGACGAGTCTTGT 2619

Mon Aug 9 13:26:08 1999

us-09-049-696-18.rng

Page 8

Db 80580 CATTGATATGTTGATTAATAAACAAATTTTAAAAAATCTCAGAGATTCAGTTAT 80637

Search completed: August 6, 1999, 04:09:06  
Job time: 1898 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:23:20 : Search time 3059.95 Seconds  
(without alignments)  
1922.929 Million cell updates/sec

Title: US-09-049-696-20

Perfect score: 2983  
Sequence: 1 GAATCAGCAGGAGATGTC.....AAATCTAACACACTGGGA 2983

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	600	20.1	734	49	A1660957	A1660957 wF20d08.x
C 2	472	15.8	716	49	A1660234	A1660234 wE68g02.x
C 3	450	15.1	490	34	AA508854	AA508854 n122h10.s
C 4	394	13.2	501	51	A1721275	A1721275 as82h08.x
C 5	338	11.3	480	35	AA581198	AA581198 nd86c07.r
C 6	304	10.2	378	31	AA296655	AA296655 EST112726
C 7	285	9.6	335	51	A1721121	A1721121 as73d08.x
C 8	213	7.1	392	50	A1687981	A1687981 wa77f06.x
C 9	182	6.1	310	31	AA297150	AA297150 EST112734
C 10	164	5.5	255	20	D25727	D25727 HDG504094
C 11	144	3.8	161	48	A1582072	A1582072 ar96a11.x
C 12	127	3.1	428	30	Z84017	Z84017 SS284017.Po
C 13	126	0.9	297	49	A1641931	A1641931 vq50e03.y
C 14	126	0.9	455	39	AA871197	AA871197 vq32c11.r
C 15	123	0.8	431	30	AA238284	AA238284 MY34g09.r
C 16	123	0.8	418	35	AA596289	AA596289 v026a08.r
C 17	123	0.8	552	37	AA688953	AA688953 vs04h07.r
C 18	123	0.8	466	37	AA691335	AA691335 vs14h06.r
C 19	123	0.8	421	37	AA691586	AA691586 vs12g06.r
C 20	123	0.8	386	37	AA692521	AA692521 vt55g08.r
C 21	123	0.8	482	37	AA711228	AA711228 vt70c11.r
C 22	123	0.8	488	37	AA733415	AA733415 vt75b06.r
C 23	123	0.8	388	37	AA734161	AA734161 vs19g06.r
C 24	123	0.8	463	39	AA839323	AA839323 v040f07.r
C 25	122	0.7	156	37	AA692444	AA692444 vt55d01.r
C 26	122	0.7	151	37	AA733390	AA733390 vt74e06.r
C 27	120	0.7	430	25	N70553	N70553 za83e12.s1
C 28	120	0.7	420	32	AA342991	AA342991 EST48700
C 29	120	0.7	388	37	AA714919	AA714919 nv49b03.r
C 30	120	0.7	574	40	AA944027	AA944027 EST199526
C 31	120	0.7	673	40	AA944028	AA944028 EST199527
C 32	120	0.7	359	41	A1047469	A1047469 ud66b12.y
C 33	119	0.6	248	20	T12464	T12464 B029F Heart
C 34	119	0.6	557	22	H10546	H10546 ym04d01.r1
C 35	119	0.6	489	22	R72678	R72678 y193d01.r1
C 36	119	0.6	295	25	D79803	D79803 HDN329B03B
C 37	119	0.6	205	25	W11372	W11372 ma86e02.r1
C 38	119	0.6	397	25	W16532	W16532 zb10b12.r1
C 39	119	0.6	479	26	W65383	W65383 zd32e05.r1
C 40	119	0.6	438	28	AA058647	AA058647 zk68g06.r
C 41	119	0.6	487	29	AA148794	AA148794 z104g01.r
C 42	119	0.6	465	30	AA240794	AA240794 mw17c03.r
C 43	119	0.6	516	30	AA255386	AA255386 SW31CA242
C 44	119	0.6	274	31	AA298669	AA298669 EST114502
C 45	119	0.6	293	31	AA321170	AA321170 EST23647

ALIGNMENTS

RESULT 1  
LOCUS A1660957/c  
DEFINITION wF20d08.x1 Soares-Dieckgraefe\_colon\_NHUC Homo sapiens CDNA clone  
IMAGE:2151151 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA  
ACCESSION A1660957  
NID 94764540



Query Match 15.8%; Score 472; DB 49; Length 716;  
 Best Local Similarity 99.8%; Pred. No. 4.9e-226;  
 Matches 522; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2461 GCTTCATCCCAAGAGCACTGAGAGAGCTTTTGTAAACGAAACAT 2520  
 DB 560 GCTCTATCCCAAGAGCACTGAGAGAGCTTTTGTAAACGAAACAT 501  
 QY 2521 ACTTTGAAATGGCAGACATCTTTTCATTCGATGAGCTGTGTAAGTCGATCG 2580  
 DB 500 ACTTTGAAATGGCAGACATCTTTTCATTCGATGAGCTGTGTAAGTCGATCG 441  
 QY 2581 AAATCAAAATATCCACATTCGACGAGATCTTTTGTATCTCCACAGACTCCGCCA 2640  
 DB 440 AAATCAAAATATCCACATTCGACGAGATCTTTTGTATCTCCACAGACTCCGCCA 381  
 QY 2641 GAGACACCTAGCTCTGATGAAGCTGCTCTCTCTAATATTCATATCAACAGACACC 2700  
 DB 380 GAGACACCTAGCTCTGATGAAGCTGCTCTCTCTAATATTCATATCAACAGACACC 321  
 QY 2701 ATTCCTGGCATTCACATTTTAAATATGTAAGTGAAGTGAAGAGAGAGCTGCTGCA 2760  
 DB 320 ATTCCTGGCATTCACATTTTAAATATGTAAGTGAAGTGAAGAGAGAGCTGCTGCA 261  
 QY 2761 ATGACCTAGGCTGATTTTGTGATGATTAATTAATCATTCCTTTTGTGA 2820  
 DB 260 ATGACCTAGGCTGATTTTGTGATGATTAATTAATCATTCCTTTTGTGA 201  
 QY 2821 TTTTAAATTTTCTAAATGATTTTGTGATGATTTTGTGATGATTTTGTGAT 2880  
 DB 200 TTTTAAATTTTCTAAATGATTTTGTGATGATTTTGTGATGATTTTGTGAT 141  
 QY 2881 ATAGTACATTTTAAATGATTTTGTGATGATTTTGTGATGATTTTGTGAT 2940  
 DB 140 ATAGTACATTTTAAATGATTTTGTGATGATTTTGTGATGATTTTGTGAT 81  
 QY 2941 TTCTGTAGGGGGGATTAATTAATGCTAAACACTGGGTA 2993  
 DB 80 TTCTGTAGGGGGGATTAATTAATGCTAAACACTGGGTA 38  
 RESULT 3  
 AA508854/c 490 bp mRNA EST 18-AUG-1997  
 LOCUS n122110.s1 NCI\_CGAP\_Co4 Homo sapiens cDNA clone IMAGE:968803 3'  
 DEFINITION mRNA sequence.  
 ACCESSION AA508854  
 NID 92246357  
 VERSION AA508854.1 GI:2246357  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 490)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:1397407.  
 JOURNAL COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,  
 M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www.bio.lnlnl.gov/dbip/image/image.html

Insert Length: 856 Std Error: 0.00  
 Seq. Primer: -40m13 fwd. ET from Amerham  
 High quality sequence stop: 404.  
 Location/Qualifiers  
 1. 490  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:968803"  
 /clone\_lib="NCI\_CGAP\_Co4"  
 /sex="pooled"  
 /tissue\_type="colon"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from pooled colon  
 tumor tissue, and was then primed with a Not I - Oligo(dT)  
 primer. Double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 This library is not normalized. Library constructed by  
 Bento Soares and M. Fatima Bonaldo."

BASE COUNT 155 a 82 c 89 g 164 t  
 ORIGIN

Query Match 15.1%; Score 450; DB 34; Length 490;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-224;  
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2534 GCACAGATCTTTTCATGATGATGATGATGATGATGATGATGATGATGAT 2593  
 DB 490 GCACAGATCTTTTCATGATGATGATGATGATGATGATGATGATGATGAT 431  
 QY 2594 CCAACATTCGACAGATGATCTTTTGTATTCCTCCACAGACTCCGCCAGACACTAGTC 2653  
 DB 430 CCAACATTCGACAGATGATCTTTTGTATTCCTCCACAGACTCCGCCAGACACTAGTC 371  
 QY 2654 CTGATGAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2713  
 DB 370 CTGATGAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311  
 QY 2714 ACATTTTAAATTAATGATGATGATGATGATGATGATGATGATGATGAT 2773  
 DB 310 ACATTTTAAATTAATGATGATGATGATGATGATGATGATGATGATGAT 251  
 QY 2774 GAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2833  
 DB 250 GAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 191  
 QY 2834 TAAATGATTTTGAAGCTTCTGATGAGGGGCGATATGATGATGATGATGAT 2893  
 DB 190 TAAATGATTTTGAAGCTTCTGATGAGGGGCGATATGATGATGATGATGAT 131  
 QY 2894 ACTAATGATTTCTGATGAGGGGCGATATGATGATGATGATGATGATGATGAT 2953  
 DB 130 ACTAATGATTTCTGATGAGGGGCGATATGATGATGATGATGATGATGATGAT 71  
 QY 2954 CGATTAATTAATTAATGCTTAACCACTGGGTA 2983  
 DB 70 CGATTAATTAATTAATGCTTAACCACTGGGTA 41

RESULT 4  
 AI721275/c 501 bp mRNA EST 10-JUN-1999  
 LOCUS as82N08.x1 Barstead colon HPLR7 Homo sapiens cDNA clone  
 DEFINITION IMAGE:2335263 3' similar to SW:ECIC\_BOVIN P54281 EPIHELIAL  
 CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.  
 ACCESSION AI721275  
 NID 95038531  
 VERSION AI721275.1 GI:5038531  
 KEYWORDS EST.  
 SOURCE human.









TITLE Raymond, L., Wei, Y.F., Ming, J., Xu, C., Yu, G.L., Ruben, S.M.,  
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,  
Fraser, C.M. and Venter, J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)  
MEDLINE 96026280  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1288566.  
Other\_ESTs: TRC167738  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/ngi/ngi.html>)  
Seq primer: M13 Reverse

FEATURES  
Source Location/Qualifiers  
1..310

/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):127942"  
/db\_xref="taxon:9606"  
/clone\_1lb="Colon 1"  
/dev\_stage="adult"  
/note="Organ: colon; Vector: pBluescript SK-; Site\_1:  
ECORI; Site\_2: XhoI"  
BASE COUNT 97 a 61 c 58 g 91 t 3 others  
ORIGIN

Query Match 6.1%; Score 182; DB 31; Length 310;  
Best Local Similarity 99.3%; Pred. No. 2,9e-84;  
Matches 282; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2306 CGGAATTCACGGGGGAGCTCTAATGACTTGACAGCTCCCGGGATGATATG 2365  
|||||  
Db 1 CGGAATTCACGGGGGAGCTCTAATGACTTGACAGCTCCCGGGATGATATG 60  
QY 2366 ACCATGAGACGCTCACAGATATATCATTCGATATGATACAGATTTCTGATCAGAG 2425  
|||||  
Db 61 ACCATGAGACGCTCACAGATATATCATTCGATATGATACAGATTTCTGATCAGAG 120  
QY 2426 ACAAGTTCATGAATCTCTCAAGTGAATACTGCTCTCATCCCAAGGACCAACT 2485  
|||||  
Db 121 ACAAGTTCATGAATCTCTCAAGTGAATACTGCTCTCATCCCAAGGACCAACT 180  
QY 2486 CTGAGGAGCTTTTCTTTTAAACCAAAAACATTACTTTGAAAATGGCAGATCTTT 2545  
|||||  
Db 181 CTGAGGAGCTTTTCTTTTAAACCAAAAACATTACTTTGAAAATGGCAGATCTTT 240  
QY 2546 TCATTGCTATTCAGGCTGTGATAGGTCGATCGAATCAGAA 2589  
|||||  
Db 241 TCATTGCTATTCAGGCTGTGATAGGTCGATCGAATCAGAA 284

RESULT 10  
D25727 255 bp mRNA EST 30-NOV-1995  
LOCUS HUMGS04094 Human colon mucosa Homo sapiens cDNA clone cm2037 3',  
DEFINITION mRNA sequence.  
ACCESSION D25727  
NID 9500422  
VERSION D25727.1 GI:500422  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 255)  
AUTHORS Okubo, K., Yoshii, J., Yokouchi, H., Kameyama, M. and Matsubara, K.

TITLE Global analysis of gene expression in colon mucosa: a large scale  
random cDNA sequencing analysis  
JOURNAL Unpublished (1994)  
COMMENT

Contact: Okubo, K., Itoh, K., Yoshii, J., Yokouchi, H. and Matsubara, K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.

FEATURES  
Source Location/Qualifiers  
1..255

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cm2037"  
/clone\_1lb="Human colon mucosa"  
/note="Adult male, tissue type = colon mucosa"  
BASE COUNT 76 a 52 c 39 g 80 t 8 others  
ORIGIN

Query Match 5.5%; Score 164; DB 20; Length 255;  
Best Local Similarity 100.0%; Pred. No. 7.8e-75;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2623 CCTCCACAGACTCGCCAGAGACCTAGCTGATGAACGTCCTGTCCTAAT 2682  
|||||  
Db 49 CCTCCACAGACTCGCCAGAGACCTAGCTGATGAACGTCCTGTCCTAAT 108  
QY 2683 ATTCATATCAACAGACACCATTCCTGCGATTCACATTTTAAATTAATGATGAGATGATA 2742  
|||||  
Db 109 ATTCATATCAACAGACACCATTCCTGCGATTCACATTTTAAATTAATGATGAGATGATA 168  
QY 2743 GGAGACTGCAAGCTGTCATATGCTAGGCTGAATTTTTCAG 2786  
|||||  
Db 169 GGAGACTGCAAGCTGTCATATGCTAGGCTGAATTTTTCAG 212

RESULT 11  
A1582072/c 161 bp mRNA EST 06-APR-1999  
LOCUS a196a11.x1 Bartshead colon HPLRB7 Homo sapiens cDNA clone  
DEFINITION IMAGE:2173436 3', mRNA sequence.  
ACCESSION A1582072  
NID 94567969  
VERSION A1582072.1 GI:4567969  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 161)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubugue, T., Gelsel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F.,  
Therling, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
TITLE On Mar 19, 1997 this sequence version replaced gi:1900677.  
JOURNAL  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40UP from Gluco  
High quality sequence stop: 115.

FEATURES  
Source Location/Qualifiers  
1..161  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2173436"

```

/clone_11b-"Barstead colon HPLR87"
/dev="male"
/sex="male"
/dev_stage="adult", age 25"
/lab_host="DH10B (phage resistant)"
/notes="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTGCAGATCTGGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[5' AATTCAGTACGAT 3' and 5' ATTCAGTG 3'], digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pT73 vector. Library constructed by Bob
Barstead."

```

[illegible]

RESULT	12
LOCUS	284017
DEFINITION	S284017 Porcine small intestine cDNA library Sus scrofa clone c12a03 5', mRNA sequence.
ACCESSION	284017
NID	G1806336
VERSION	284017.1
KEYWORDS	GI:1806336
SOURCE	EST.
ORGANISM	Pig. <i>Sus scrofa</i>
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus. 1 (bases 1 to 428)
AUTHORS	Wintero,A.R., Freholm,M. and Davies,W.
TITLE	Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones
JOURNAL	Mamm.Genome 7 (7), 509-517 (1996)
MEDLINE	96327607
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1104772.

**CONTACT:** A.K. Winterøe  
 Department of Animal Science and Animal Health, Division of Animal  
 Genetics, The Royal Veterinary and Agricultural University  
 Bulowsovej 13, 1870 Frederiksberg C, Denmark.

**FEATURES**  
 Location/Qualifiers  
 1..428  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone="c12a03"  
 /clone\_lib="Porcine small intestine cDNA library"  
 /note="directionally cloned cDNA in X1-blue MRF"  
 BASE COUNT 134 a 87 c 100 g 102 t 5 others  
 ORIGIN

Query Match	1.0%	Score 31:	DB 30:	Length 428:
Best Local Similarity	100.0%	Pred. No. 2	7e-05:	
Matches	31:	Conservative	0:	Mismatches 0:
				Indels 0:
				Gaps 0:
QY	241	TATTTCAAAAATGTGCGCATTTTGATTCGTG	271	
db	242	TATTTCAAAAATGTGCGCATTTTGATTCCTG	272	

RESULT 13  
A1641931  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

297 bp mRNA EST 29-APR-1999  
V950603.V1 Barstead bowel MRLB93 Mus musculus cDNA clone  
IMAGE:1097692 5' similar to TR:088826 088826 GOB-5 PROTEIN. ; mRNA  
sequence.  
A1641931  
A16420406  
A1641931.1 GI:4720406  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eukaryota; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 297)  
Warr,M., Hillier,L., Kucba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Thaisang,M., Page,D., Harvey,N., Schurk,R.,  
Person,B., Swaller,T., Gibbons,M., Pape,D., Harney,N., Schurk,R.,  
Ritter,E., Kohn,S., Shin,T., Jackson,T., Cardenas,M., McCann,R.,  
Waterson,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
On Mar 10, 1998 this sequence replaced gi:2948705.

Contact: Marra M/Mashu-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MG1603924  
 This read is a RESSEQUENCE of a previously sequenced mouse clone  
 This read has been verified (found to hit its original self in the  
 correct orientation)  
 Putative full length read  
 vector to vector length is 481  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Glibco  
 High quality sequence stop: 288.  
 location/Qualifiers  
 1..297  
 source  
 1..297  
 "Mus musculus"

	FEATURES	SOURCE
location/Qualifiers	1..297	
/organism="Mus musculus"		
/strain="FVB/N"		
/db_xref="taxon:10090"		
/clone="IMAGE:1097692"		
/clone_1kb="Barstead bowel MPLB9"		
/tissue_type="bowel"		
/dev_stage="8 weeks"		
/lab_host="DH10B"		
/note-Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5'] TGTACGAAATCTGAAGTGAGCGAGCGCCGCCCTTTTTTTTTTTTTTTT		
3']; double-stranded cDNA was ligated to Eco RI adaptors (AATTGGCATCCTTG), digested with Not I and cloned into the		
Not I and Eco RI sites of the modified pT73 vector.		
Source irradiated bowel harvested 72 hours after		
irradiation (1400 Gys). Library constructed by Bob		
Barstead."		

Query Match	0.98;	Score 27;	DB 49;	Length 297;
Best Local Similarity	100.0%;	Prod. No. 0.0035;		
Matches 27;	Conservative 0;	Indels 0;	Gaps 0;	
QY 2689	ATCAACGACCACTTCCTGGCATTTAC	2715		



This Page Blank (uspto)

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:23:08 ; Search time 3059.95 Seconds  
(without alignments)  
1813.342 Million cell updates/sec

Title: US-09-049-696-18

Perfect score: 2813  
Sequence: 1 GAAATCAGGAGGAGATGTAC.....AAATAATCATTCATCCTTA 2813

Scoring table: OLIGO\_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	598	21.3	734	49	A1660957	A1660957 wf20d08.x
C 2	394	14.0	501	51	A1721275	A1721275 as82h08.x
C 3	338	12.0	480	35	AA581198	AA581198 nd38c07.x
C 4	304	10.8	378	31	AA296955	AA296955 EST112726
C 5	301	10.7	716	49	A1660234	A1660234 we68g02.x
C 6	285	10.1	335	51	A1721121	A1721121 as73d08.x
C 7	279	9.9	490	34	AA508854	AA508854 n122h10.s
C 8	182	6.5	310	31	AA297150	AA297150 EST112734
C 9	180	6.4	392	50	A1687981	A1687981 wa77f06.x
C 10	164	5.8	255	20	D25727	D25727 HUMG504094
C 11	98	3.5	161	48	A1582072	A1582072 at96a11.x
C 12	31	1.1	428	30	Z84017	Z84017 SS284017.Po
C 13	27	1.0	297	49	A1641931	A1641931 vq30e03.y
C 14	26	0.9	455	39	AA871197	AA871197 vq32c11.y
C 15	23	0.8	431	30	AA238284	AA238284 my34g09.x
C 16	23	0.8	418	35	AA596289	AA596289 v026a08.x
C 17	23	0.8	552	37	AA688953	AA688953 v04b07.x
C 18	23	0.8	646	37	AA691335	AA691335 vs14h06.x
C 19	23	0.8	421	37	AA691586	AA691586 vs12g06.x
C 20	23	0.8	386	37	AA692521	AA692521 vt59g08.x
C 21	23	0.8	482	37	AA711228	AA711228 vt70c11.x
C 22	23	0.8	488	37	AA733415	AA733415 vt75b06.x
C 23	23	0.8	388	37	AA734161	AA734161 vs19g06.x
C 24	23	0.8	463	39	AA839323	AA839323 v04f07.x
C 25	20	0.7	430	25	N70553	N70553 za83e12.s1
C 26	20	0.7	420	32	AA342991	AA342991 EST48700
C 27	20	0.7	388	37	AA714919	AA714919 nv49b03.x
C 28	20	0.7	574	40	AA944027	AA944027 EST199526
C 29	20	0.7	673	40	AA944028	AA944028 EST199527
C 30	20	0.7	359	41	A1047469	A1047469 ud66b12.y
C 31	19	0.7	248	20	T12464	T12464 B029F Heart
C 32	19	0.7	557	22	H10546	H10546 ym04d01.r1
C 33	19	0.7	489	22	R72678	R72678 v193d01.r1
C 34	19	0.7	295	25	D79803	D79803 HUM329B03B
C 35	19	0.7	205	25	W11372	W11372 m886e02.r1
C 36	19	0.7	397	25	W16332	W16332 zb10b12.r1
C 37	19	0.7	479	26	W65383	W65383 zd32e05.r1
C 38	19	0.7	438	28	AA058647	AA058647 zk69g06.x
C 39	19	0.7	487	29	AA148794	AA148794 z104g01.x
C 40	19	0.7	465	30	AA240794	AA240794 mw17c03.r
C 41	19	0.7	516	30	AA255386	AA255386 SW31CA242
C 42	19	0.7	274	31	AA298669	AA298669 EST114502
C 43	19	0.7	293	31	AA321170	AA321170 EST23647
C 44	19	0.7	349	32	AA376662	AA376662 EST89286
C 45	19	0.7	415	37	AA669224	AA669224 ac16b04.s

ALIGNMENTS

RESULT 1  
LOCUS A1660957 734 bp mRNA EST 10-MAY-1999  
DEFINITION wf20d08.x1 Soares\_Dieckgraefe\_colon\_NHUC Homo sapiens cDNA clone  
IMAGE:2351151 3' similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA  
sequence.  
ACCESSION A1660957  
NID g4764540

VERSION A1660957.1 GI:4764540  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 734)  
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 TITLE Unpublished (1997)  
 JOURNAL On Mar 20, 1998 this sequence version replaced gi:2980033.  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40UP from GIBCO  
 High quality sequence stop: 477.  
 Location/Qualifiers  
 1..734  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="X"  
 /clone="IMAGE:2351151"  
 /clone\_lib="Soares\_Dieckgraefe\_colon\_NHUC"  
 /tissue\_type="colonic mucosa from 5 ulcerative colitis  
 patients"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: colon; Vector: pTR73-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTACCACTGAGAGGAGCGGCGCTGCTTCTGATGCCAATCTCCATCTGATCTC  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pTR73 vector. Library  
 went through one round of normalization. Tissue samples  
 provided by Dr. Brian Dieckgraefe (Washington University,  
 dieck@im.wustl.edu); colonic mucosa represents a range of  
 disease involvement from mild cryptitis to severe  
 ulceration, fibrosis, and degeneration. Library  
 constructed by Bento Soares and M. Fatima Bonaldo. "  
 BASE COUNT 200 a 143 c 159 g 228 t 4 others  
 ORIGIN  
 Query Match 21.3%; Score 598; DB 49; Length 734;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 2215 ACATCTCTGGGAGGCTCATTTGTGCTTCTGATGCCAATCTCCATCTGATCTC 2274  
 |||||||  
 Db 602 ACATCTCTGGGAGGCTCATTTGTGCTTCTGATGCCAATCTCCATCTGATCTC 543  
 Oy 2275 TTCCCACTGGGCAATACCGACGACGGAAGGGAATTCAGGGGGAGTCCATTAT 2334  
 |||||||  
 Db 542 TTCCCACTGGGCAATACCGACGACGGAAGGGAATTCAGGGGGAGTCCATTAT 483  
 Oy 2335 CTGACTTGGACAGCTCTCTGGGATGATTTATGACCATGGAACAGCTCAAGTATCAT 2394  
 |||||||  
 Db 482 CTGACTTGGACAGCTCTCTGGGATGATTTATGACCATGGAACAGCTCAAGTATCAT 423  
 Oy 2395 CGAATAGTACAGTATCTTGTGATCTCAGAGACAAGTTCATGAATCTCTCAAGTAA 2454  
 |||||||  
 Db 422 CGAATAGTACAGTATCTTGTGATCTCAGAGACAAGTTCATGAATCTCTCAAGTAA 363  
 Oy 2455 ACTACTGCTCTATCCCAAGGAAGCAACTCTGAGGAATCTTTTGTGTTAAACAGAA 2514  
 |||||||  
 Db 362 ACTACTGCTCTATCCCAAGGAAGCAACTCTGAGGAATCTTTTGTGTTAAACAGAA 303  
 Oy 2515 AACATCTCTTGAATAATGGACAGATCTTTTATCTGATGAGCTGTGTAAGTTC 2574  
 |||||||

Db 302 AACATCTCTTGAATAATGGACAGATCTTTTATCTGATGAGCTGTGTAAGTTC 243  
 Oy 2575 GATCTGAATTCAGAAATATCCACATTCGACAGATCTTTGTTATCTCTCCACAGACT 2634  
 |||||||  
 Db 242 GATCTGAATTCAGAAATATCCACATTCGACAGATCTTTGTTATCTCTCCACAGACT 183  
 Oy 2635 CCGCAGAGACACCTAGTCCGATGAAGACGTCCTCTGCTCTTAATTCATATCAAC 2694  
 |||||||  
 Db 182 CCGCAGAGACACCTAGTCCGATGAAGACGTCCTCTGCTCTTAATTCATATCAAC 123  
 Oy 2695 AGCACATCTCTGGCATTACATTTTAAATATATGTAAGTATGAGAGAACTGCAG 2754  
 |||||||  
 Db 122 AGCACATCTCTGGCATTACATTTTAAATATATGTAAGTATGAGAGAACTGCAG 63  
 Oy 2755 CTGTCAATAGCTTAGGCGCTGAATTTTGTGATGAATAAATAATCATTCATCTT 2812  
 |||||||  
 Db 62 CTGTCAATAGCTTAGGCGCTGAATTTTGTGATGAATAAATAATCATTCATCTT 5  
 RESULT 2  
 A1721275/c 501 bp mRNA EST 10-JUN-1999  
 LOCUS  
 DEFINITION aas2h08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
 IMAGE:235263 3' similar to SW:BCLC\_BOVIN P54281 EPTHELIAL  
 CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.  
 A1721275  
 ACCESSION 95038531  
 NID A1721273.1 GI:5038531  
 VERSION EST.  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 501)  
 Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maria,M.,  
 Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
 Theising,B., White,Y., Wylie,T., Waterson,R. and Wilson,R.  
 Mashu-NCI human EST Project  
 TITLE Unpublished (1997)  
 JOURNAL On Jun 22, 1998 this sequence version replaced gi:3246918.  
 COMMENT  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40UP from GIBCO  
 High quality sequence stop: 394.  
 Location/Qualifiers  
 1..501  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="Xp11.2"  
 /clone="IMAGE:235263"  
 /clone\_lib="Barstead colon HPLRB7"  
 /sex="male"  
 /dev\_stage="adult, age 25"  
 /lab\_host="DH10B (phage resistant)"  
 /note="Organ: colon; Vector: pTR73-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTACCACTGAGAGGAGCGGCGCTTCTGATGCCAATCTCCATCTGATCTC  
 double-stranded cDNA was ligated to Eco RI adaptors  
 [5', AATCACTAGTAT 3' and 5' AATCACTAGT 3'], digested  
 with Not I and cloned into the Not I and Eco RI sites of  
 the modified pTR73 vector. Library constructed by Bob  
 Barstead."

BASE COUNT

158 a 83 c 106 g 154 t



kelley J.M., Kelley J.C., Liu L.-I., Marmaros S.M., Merrick J.M.,  
Moreno-Palancques R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,  
Phillips C.A., Ryder S.E., Scott J.L., Saudak D.M., Shirley R.,  
Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y.,  
Bednarik D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,  
Dimke D., Feng D.-F., Ferite A., Fischer C., Hastings G.A.,  
He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K.,  
Kozak D.L., Kunsch C., Hungjun J., Li H., Meisner P.S., Olsen H.,  
Raymond L., Wei Y.F., Wang J., Xu C., Yu G.L., Ruden S.M.,  
Dillon P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C.,  
Fraser C.M. and Venter J.C.

TITLE Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence

JOURNAL Nature 377 (6547 Suppl.), 3-174 (1995)

MEDLINE 96036280

COMMENT On Sep 12, 1996 this sequence version replaced g1:1288227.  
g1:1288227

**Source**

Location/Qualifiers	1. .378
---------------------	---------

BASE COUNT	112 a	79 c	64 g	119 t	4 others
ORIGIN					

```

Query Match      10.8%   Score 304   DB 31;   Length 378;
Best Local Similarity 100.0%   Pred. No. 3.6e-151;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

Oy	2461	6CTGTCATCCCAAGGAAGCCAACTGTGAGGAAGTCCTTTTGTTTAAACCAAGAAACCTT	2520
Db	19	GGCTGTCATCCCAAGGAAGCCAACTGTGAGGAAGTCCTTTTGTTTAAACCAAGAAACCTT	78
Oy	2521	ACTTTTGAAGAAATGGCACAGATCTTTTCATTGCTATTCAGGCTGTGATTAAGTCGATCTG	2580
Db	79	ACTTTTGAAGAAATGGCACAGATCTTTTCATTGCTATTCAGGCTGTGATTAAGTCGATCTG	138
Oy	2581	AATATGAAATATCCACATTCGACAGATTACTTGTATTCTCTGCACAGACTCCGCCA	2640
Db	139	AATATGAAATATCCACATTCGACAGATTACTTGTATTCTCTGCACAGACTCCGCCA	198
Oy	2641	GAGACACCTAGTCGTGATGAAAGCTGTGCTCTTGTCTTAATTTGATATCAAGAGCAAC	2700
Db	199	GAGACACCTAGTCGTGATGAAAGCTGTGCTCTTGTCTTAATTTGATATCAAGAGCAAC	258
Oy	2701	AATCCGCGATTCACATTTTAAAAATTTATGTGAAGTGATAGAGAACTGCAGCTGTCA	2760
Db	259	AATCCGCGATTCACATTTTAAAAATTTATGTGAAGTGATAGAGAACTGCAGCTGTCA	318
Oy	2761	ATAG 2764	
Db	319	ATAG 322	

RESULT	5	
AI660234/c		
LOCUS	AI660234	716 bp mRNA EST
		10-MAY-1999

## DEFINITIO

DEFINITION	IMAGE	SEQUENCE
we68g02.x1 Soares_Dieckgraefe.COLON.NHCD Homo sapiens cDNA clone IMAGE:2340290 3' similar to TR:088826 088826 GOB-5 PROTEIN. ;, mRNA		

ACCESSION

NID	94703804
VERSION	AI660234.1
	GI:4763804

**KEYWORDS**  
**SOURCE**

## ORGANIS

## REFERENCE AUTHORS

**TITLE**

**JOURNAL  
COMMENT**

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: ~400P from Glibco  
High quality sequence stop: 387.

FEATURES  
source

BASE COUNT	220 a	123 c	140 g	233 t
ORIGIN				

Query Match	10.7%	Score 301;	DB 49;	Length 716;
Similarity	99.7%	Pred. No. 1.3e+149;		
Matches 351; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy	2461	GCTCTCATCCCAAGGAAGCCAACTCTGAGGAAGCTCTTTTGTTTAAACACAAACAACTT	2520
Db	560	GCTCTATCTCCCAAGGAAGCCAACTCTGAGGAAGCTCTTTTGTTTAAACCAAAACCACTT	501
Qy	2521	ACTTTGAAAATGAGCAGACAGATCTTTTATTGCTATTACAGGCTGTGTAAAGTCGATCTG	2580
Db	500	ACTTTGAAAATGGCAGACAGATCTTTTATTGCTATTACAGGCTGTGTAAAGTCGATCTG	441
Qy	2581	AAATCAGAAATTTCCAAACATTGACAGAGTATCTTGTTTATCTCCACAGACTCGGCA	2640
Db	440	AAATCAGAAATTTCCAAACATTGACAGAGTATCTTGTTTATCTCCACAGACTCGGCA	381
Qy	2641	GAGACACTAGTCTCGATGAAGAAAGCTGCTCTCTTCTTAATATTCATATCAACAGCAC	2700
Db	380	GAGACACTAGTCTCGATGAAGAAAGCTGCTCTCTCTTAATATTCATATCAACAGCAC	321
Qy	2701	ATTCTCGCAATTCACATTTTAAAAATTTGTGGAAAGGATAGAGAACTGCAGGCTGCA	2760



|||||  
Db 320 ATTCTGCGATCATTCTTTAAATATATGAGAGTGATGAGAGAACTGACCTGCA 261  
|||  
QY 2761 ATAGCGTAGGCGTGAATTTTGTGAGATAAATAATATCATTCCTT 2812  
|||||  
Db 260 ATAGCGTAGGCGTGAATTTTGTGAGATAAATAATATCATTCCTT 209  
|||  
RESULT 6  
AI721121/c 335 bp mRNA EST 10-JUN-1999  
LOCUS as73608.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
DEFINITION IMAGE:2334351 3' similar to TR:088826 088826 G08-5 PROTEIN. ; mRNA  
SEQUENCE.  
AI721121  
ACCESSION AI721121 GI:5038377  
NID 95038377  
VERSION AI721121.1 GI:5038377  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 335)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI Human EST Project  
Unpublished (1997)  
On Jun 22, 1998 this sequence version replaced gi:3246762.  
TITLE  
JOURNAL  
COMMENT  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estevason.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from Gibco  
High quality sequence stop: 312.  
Location/Qualifiers  
1. 335  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2334351"  
/clone\_1lb="Barstead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGATGAGAGTCTTTTGTGTTAAACGAGAAACATTCATTGTAATGCGACAGATCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[5' AATTCCTGATGAT 3' and 5' ATTACTAGT 3'], digested  
with Not I and cloned into the Not I and Eco RI sites of  
the modified pT73 vector. Library constructed by Bob  
Barstead."

BASE COUNT 98 a 57 c 71 g 109 t  
ORIGIN  
Query Match 10.1%; Score 285; DB 51; Length 335;  
Best Local Similarity 100.0%; Pred. No. 5.1e-141;  
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2487 TGAGGAGTCTTTTGTGTTAAACGAGAAACATTCATTGTAATGCGACAGATCTTTT 2546  
|||||  
Db 324 TGAGGAGTCTTTTGTGTTAAACGAGAAACATTCATTGTAATGCGACAGATCTTTT 265  
|||||  
QY 2547 CATGCTATTCAGGCTGTGATAGTGATGTAATGCAAAATATCACAATTCGACG 2606  
|||||

|||||  
Db 264 CATGCTATTCAGGCTGTGATAGTGATGTAATGCAAAATATCACAATTCGACG 205  
|||||  
QY 2607 AGATCTTTGTTATTCCTCCACAGACTCCGCGAGAGACACTAGTCTGTGTAACGTC 2666  
|||||  
Db 204 AGTATCTTTGTTATTCCTCCACAGACTCCGCGAGAGACACTAGTCTGTGTAACGTC 145  
|||||  
QY 2667 TGCTCTTGTCTTAATATTCATATCAACAGACACCATTCCTGGCATTTTAAAAAT 2726  
|||||  
Db 144 TGCTCTTGTCTTAATATTCATATCAACAGACACCATTCCTGGCATTTTAAAAAT 85  
|||||  
QY 2727 TATGTGAGAGTGATAGAGAGTGCAGTGTCTCAATAGCTAGG 2771  
|||||  
Db 84 TATGTGAGAGTGATAGAGAGTGCAGTGTCTCAATAGCTAGG 40  
|||||  
RESULT 7  
AA508854/c 490 bp mRNA EST 18-AUG-1997  
LOCUS AA508854  
DEFINITION n122h10.s1 NCI\_CGAP\_C04 Homo sapiens cDNA clone IMAGE:968803 3',  
mRNA sequence.  
ACCESSION AA508854  
NID 92246357  
VERSION AA508854.1 GI:2246357  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 490)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:3397407.  
TITLE  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.llnl.gov/bdrp/image/image.html  
Insert Length: 856 Std Error: 0.00  
Seq primer: -40m13 fwd. RT from Amersham  
High quality sequence stop: 404.  
Location/Qualifiers  
1. 490  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:968803"  
/clone\_1lb="NCI\_CGAP\_C04"  
/sex="pooled"  
/tissue="colon"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from pooled colon  
tumor tissue, and was then primed with a Not I - oligo(dT)  
primer. Double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
This library is not normalized. Library constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 155 a 82 c 89 g 164 t  
ORIGIN  
Query Match 9.9%; Score 279; DB 34; Length 490;

Best Local Similarity 100.0%; Pred. No. 7.7e-138; Mismatches 279; Conservative 0; Indels 0; Gaps 0;

2534 GCACACATCTTTTCATCTGCTATTCAGCTGTTGATAGCTGATCTGAATCAGAAATAT 2593

DB 490 GCACACATCTTTTCATCTGCTATTCAGCTGTTGATAGCTGATCTGAATCAGAAATAT 431

2594 CCAACATTCGACGAGTATCTTTGTTTATTCCTCCACAGACTCCGACAGACACTAGTC 2653

DB 430 CCAACATTCGACGAGTATCTTTGTTTATTCCTCCACAGACTCCGACAGACACTAGTC 371

2654 CTATGTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2713

DB 370 CTATGTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311

2714 ACATTTTAAATATATGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 2773

DB 310 ACATTTTAAATATATGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 251

2774 GAATTTTTCATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2812

DB 250 GAATTTTTCATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 212

RESULT 8 AA297150 310 bp mRNA EST 18-APR-1997

LOCUS EST112734 Colon I Homo sapiens cDNA 5' end, mRNA sequence.

DEFINITION AA297150 91949524

KEYWORDS AA297150.1 GI:1949524

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 310)

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,

Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,

White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C.,

Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,E., Fline,L.D.,

Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.,

Glodex,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,

Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,

Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,

Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shiley,R.,

Small,K.V., Spriggs,T.A., Uteback,T.R., Weidman,J.F., Li,Y.,

Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,

Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,

He,W.W., Hu,Y.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,

Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Melsner,P.S., Olsen,H.,

Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,

Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,

Fraser,C.M., and Venter,J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

On Sep 12, 1996 this sequence version replaced gi:1288566.

Other ESTs: TRC167738

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1..310

/organism="Homo sapiens"

/db\_xref="ATCC (Inhost):127942"

/db\_xref="taxon:9606"

/clone\_lib="Colon I"

/dev\_stage="adult"

/note="organ: colon; Vector: pbluescript SK-; Site-1:

EcoRI, Site-2: XhoI"

BASE COUNT 97 a 61 c 58 g 91 t 3 others

ORIGIN

Query Match 6.5%; Score 182; DB 31; Length 310;

Best Local Similarity 99.3%; Pred. No. 4.5e-86;

Mismatches 282; Conservative 0; Indels 0; Gaps 0;

2306 CGGAATTCACGGGGGAGCTCTCATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2365

DB 1 CGGAATTCACGGGGGAGCTCTCATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60

2366 ACCATGGAACGCTCACAAGTATATCATTCGAATTAAGTACAGATCTGATCTCAGAG 2425

DB 61 ACCATGGAACGCTCACAAGTATATCATTCGAATTAAGTACAGATCTGATCTCAGAG 120

2426 ACAAGTCAATGAATCTCTTCAAGTATATCACTGCTCTCATCCCAAGAGCAACT 2485

DB 121 ACAAGTCAATGAATCTCTTCAAGTATATCACTGCTCTCATCCCAAGAGCAACT 180

2486 CTGAGGAAGCTCTTTGTTTAAACAGAAACATTACTTTGAAATAGGACAGATCTT 2545

DB 181 CTGAGGAAGCTCTTTGTTTAAACAGAAACATTACTTTGAAATAGGACAGATCTT 240

2546 TCATTGCTATTCAGGCTGTGATTAAGTGTGATCTGAAATCGAA 2589

DB 241 TCATTGCTATTCAGGCTGTGATTAAGTGTGATCTGAAATCGAA 284

RESULT 9

LOCUS A1687981 392 bp mRNA EST 27-MAY-1999

DEFINITION wa77f06.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone

IMAGE:3302211 3' similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA

sequence.

ACCESSION A1687981

VERSION 94899275

KEYWORDS A1687981.1 GI:4899275

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 392)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Mar 10, 1998 this sequence version replaced gi:2947973.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1350

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium ([info@image.jhmi.gov](mailto:info@image.jhmi.gov)) for further information.

Seq primer: -40UP from Gibco.

Location/Qualifiers

1..392

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker. Site-1: Not I; Site-2: Eco RI;

Equal amounts of plasmid DNA from three normalized





WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:602188  
Seq primer: -28m3 rev2 ET from Amersham  
High quality sequence stop: 414.

## FEATURES

source

1..455  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1095956"  
/clone\_1ib="Barstead bowel MPLRB9"  
/tissue\_type="bowel"  
/dev\_stage="8 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTACGATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[ATTCGATCCTG], digested with Not I and cloned into the  
Not I and Eco RI sites of the modified pT73 vector.  
Source irradiated bowel harvested 72 hours after  
irradiation (1400 Gys). Library constructed by Bob  
Barstead."

BASE COUNT 122 a 123 c 105 g 105 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 26; DB 39; Length 455;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1573 TTGTTCTTATCAGCTGGACACGCA 1598

DB 161 TTGTTCTTATCAGCTGGACACGCA 186

## RESULT 15

AA238284

LOCUS

DEFINITION 431 bp mRNA EST 03-MAR-1997  
clone IMAGE:69792.5' similar to TR:G1184066 G1184066  
CALCIUM-ACTIVATED CHLORIDE CHANNEL.; mRNA sequence.

ACCESSION

AA238284

NID

91862488

VERSION

AA238284.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 431)

Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Giesel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

On Sep 12, 1996 this sequence version replaced g1:1393580.

Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:431352  
Putative full length read  
vector to vector length is 439  
Seq primer: -28m3 rev2 ET from Amersham.

## FEATURES

source

1..431  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:69792"  
/clone\_1ib="Barstead mouse pooled organs MPLRB4"  
/sex="mixed"  
/tissue\_type="pooled organs"  
/dev\_stage="7 day"  
/lab\_host="DH10B"  
/note="organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACGATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[GTGGATTCGCTACC], digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
Library constructed by Bob Barstead."

BASE COUNT 120 a 99 c 98 g 114 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 23; DB 30; Length 431;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2689 ATCAGCAGCAGCTTCCTGCGAT 2711

DB 211 ATCAGCAGCAGCTTCCTGCGAT 233

Search completed: August 6, 1999, 09:23:16  
Job time: 6267 sec

This Page Blank (uspto)

---

OM nucleic - nucleic search, using sw model

28.765 Million cell updates/sec

Title: US-09-049-696-17  
Page: 106

Sequence: 1 GGCATTACATTTTAAAT...AAATAAATCATTCATCCTT 106

Scoring table: IDENTITY\_NUC

Database : N\_Geneseq\_36: \*

**Pred. NO.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	103	97.2	255	1	T22483	Human gene signatu
2	76.8	72.5	878	1	T45884	Human colon speci
3	76.8	72.5	878	1	V16672	Polynucleotide seq
4	34	32.1	110000	1	X20248_-08	Continuation (9 of
5	33.2	31.3	2520	1	O77884	Neural thred prob
6	33.2	31.3	2520	1	T27766	AD 16c human neutra
7	28	26.4	2293	1	T89279	TPR-F1 gene promo
8	28	26.4	6735	1	V52337	Streptococcus pneu
9	28	26.4	110000	1	X20248_-05	Continuation (6 of
10	27.6	26.0	134525	1	O04525	Total base sequenc
11	27.6	26.0	4436	1	N81716	Human growth hormo
12	27.6	26.0	435	1	V57245	Clone #1 from muta
13	27.4	25.8	1133	1	V20912	Human DBI/ACBP -11
14	27.4	25.8	6217	1	V65252	DNA encoding S. pn
15	27	25.5	13865	1	V40401	Human tissue factor
16	26.8	25.3	14800	1	X18205	OPF8 growth sequen
17	26.6	25.1	2676	1	V31198	E. coli j96 pathoge
18	26.4	24.9	110000	1	V21209_-15	Continuation (16 o
19	26	24.5	4434	1	N91325	DNA encoding human
20	26	24.5	1597	1	O10867	T18 oncogene. Effi
21	26	24.5	1597	1	O14048	Human OT18 clone p
22	26	24.5	4366	1	T94063	Human growth hormo
23	26	24.5	19250	1	V52168	Streptococcus pneu
24	25.8	24.3	409	1	O60252	Human brain Expres
25	25.8	24.3	365	1	T21315	Human gene signatu
26	25.8	24.3	1392	1	T72874	DNA encoding glycy
27	25.8	24.3	543	1	T84228	DNA encoding a citi
28	25.8	24.3	58407	1	V21210	Methanococcus jann
29	25.8	24.3	110000	1	V21209_-12	Continuation (13 o
30	25.6	24.2	2404	1	T33220	Oestrogen-regulate
31	25.6	24.2	774	1	T67558	H. pylori1 secreted
32	25.6	24.2	825	1	T67967	H. pylori1 secreted
33	25.6	24.2	2404	1	T99071	Partial sequence o
34	25.6	24.2	110000	1	V21209_-10	Continuation (11 o
35	25.6	24.2	954	1	V65277	DNA encoding a S.
36	25.6	24.2	2327	1	V59607	Human secreted prot
37	25.6	24.2	2047	1	X13241	Enterococcus faeca
38	25.6	24.2	1094	1	X13577	Enterococcus faeca
39	25.6	24.2	771	1	X30423	H. pylori1 secreted
40	25.4	24.0	409	1	N91158	Clai ferment conty
41	25.4	23.8	26811	1	X20253	Borrelia burgdorfer
42	25.4	23.8	110000	1	V21209_-05	Continuation (6 of
43	25.2	23.8	5162	1	V74104	Human G-protein re

Rat pheromone rece  
Continuation (17 o

## ALIGNMENTS

RESULT	1
ID	T22483
AC	T22483;
DT	22-AUG-1996 (first entry)
DE	Human gene signature HDWGSC4094.
KW	Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW	human; cloning; mapping; non-biased library; diagnosis; detection;
KW	cell typing; abnormal cell function; ss.
OS	Homo sapiens.
PN	WO9514772-A1.
PD	01-JUN-1995.
PF	11-NOV-1994; J01916.
PR	12-NOV-1993; JP-355504.
PA	(MATS/) MATSUBARA K.
PI	(OKUB/) OKUBO K.
PT	Matsubara K, Okubo K;
PT	WPI: 95-206931/27.
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g
PT	for diagnosis of abnormal cell function, by preparing cDNA that
PT	reflects relative abundance of corresp. mRNA in specific human
PT	tissues
PS	Claim 1: Page 1138-1139; 2245pp; Japanese.
CC	A single-stranded DNA (or its complementary strand or the corresp.
CC	double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC	given in T19001-T16837 and which is able to hybridise to part of
CC	human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC	sequences were obtained from 3'-directed cDNA libraries prepared
CC	from various human tissues; synthesis of cDNA was initiated from the
CC	3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC	untranslated sequence is unique to a particular mRNA species, almost
CC	all the 3'-oriented cDNAs hybridise with specific mRNAs. Each librar
CC	is constructed so as to reflect accurately the relative abundance of
CC	different mRNAs in the particular tissue from which it was derived.
CC	The appearance frequency of a given GS in a cDNA library can be
CC	determined (esp. using primers and probes derived from the GS
CC	sequences) as a means of diagnosing abnormal cell function or for
CC	recognising different cell types.
CC	Sequence 255 BP; 76 A; 52 C; 39 G; 80 T;

Query Match	97.28;	Score 103;	DB 1;	Length 255;
-------------	--------	------------	-------	-------------

Best Local Similarity 97.2%; Pred. No. 9e-24;  
Matches 103; Conservative 0; Mismatches 3; Indels 0; Gaps 0

```

OY      1  GGCATTTACATTTTAAAAAATATGTGGAGAGTGGATAGAGAAACGTGCAGCTGCATATGCC 60
Db      133 GGCATTCACATTTTAAAAAATATGTGGAGAGTGGATAGAGAACTGCAGCTGTCAATAGCC 192
OY      61  TAGGGCTGGAATTTTGTGCAGATTAATAAAAATAAATCATTCATCCTT 106
Db      193 TAGGGCTGGAATTTTGTGCAGATTAATAAAAATCAATCATCCTT 238

RESULT      2
T45884
AC      T45884 standard; cDNA; 878 BP.
DT      13-MAR-1997 (first entry)
DE      Human colon specific gene CSG5 cDNA partial clone.
KW      Colon specific gene; CSG5; colon cancer; metastasis;
KW      gene therapy; ss.
OS      Homo sapiens.
FH      key
FT      cds
          Location/Qualifiers
          2..691
          /*tag= a

```

PN W09639419-A1.  
 PD 12-DEC-1996.  
 PR 06-JUN-1995; U07289.  
 PF 06-JUN-1995; WO-U07289.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Yu G;  
 DR WPI: 97-043054/04.  
 DR P-ESDB: W06548.  
 DR Human colon specific genes and their expression products - detection of which, in non-colon tissue samples, can be used as indication of colon cancer metastasis  
 PI colon cancer metastasis  
 PS Claim 1; Fig 5; 60pp; English.  
 CC 13 cDNA clones (745880-92), most of them partial clones, correspond to human colon specific genes, designated CSG1, CSG2, etc., that are primarily expressed in tissues derived from the colon. CSG1 and CSG10 show reduced expression in colon cancer cells as compared to that in normal cells; the remaining genes are overexpressed in colon cancer. The partial cDNA sequences can be used to isolate full-length clones and genomic clones including the complete gene. CSG nucleic acids can be used to produce CSG polypeptides (see also W06545-53) in transformed host cells, as probes to detect disorders of the colon, partic. colon cancer and colon cancer metastasis, and in gene therapy.  
 CC in gene therapy.  
 CC Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 72.5%; Score 76.8; DB 1; Length 878;  
 Best Local Similarity 91.0%; Pred. No. 1.8e-15;  
 Matches 81; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GGCATTACATTTTAAATATGCGAAGTGAAGAGACTGCAGCTGCAATAGCC 60  
 DB 717 GGCATTACATTTTAAATATGCGAAGTGAAGAGACTGCAGCTGCAATAGNC 776  
 QY 61 TAGGCTGAATTTTGTGCAGTAATAA 89  
 DB 777 TAGGGGTGAATTTTGTGCGGTGAATAA 805

RESULT 3  
 ID V16672 standard; cDNA: 878 BP.  
 AC V16672;  
 DT 22-JUN-1998 (first entry)  
 DE Polynucleotide sequence of a colon-specific gene.  
 KM Colon-specific gene; probe; detection; expression; human;  
 OS diagnostic assay; colon cancer; antibody; screening; ss.  
 FH Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 2..685  
 FT /note="no stop codon given"  
 FT /tag="a  
 PN US5733748-A.  
 PD 31-MAR-1998.  
 PR 06-JUN-1995; 469667.  
 PF 06-JUN-1995; US-469667.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen C, Yu G;  
 DR WPI: 98-229823/20.  
 DR P-ESDB: W46879.  
 DR Colon-specific nucleic acids - useful as probes for detecting colon cancer micrometastases  
 PT Claim 15; Fig 5A-B; 51pp; English.  
 CC V16668-81 represent polynucleotide sequences of partial or full-length cDNA clones of colon-specific genes. The polynucleotides can be used as probes to detect expression of the corresponding human genes, e.g. in diagnostic assays for detecting micrometastases of colon cancer.  
 CC Recombinant cells containing the polynucleotides can be used to produce the polypeptides, in order that antibodies can be raised and used in further screening or diagnostics.  
 CC Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 72.5%; Score 76.8; DB 1; Length 878;  
 Best Local Similarity 91.0%; Pred. No. 1.8e-15;  
 Matches 81; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GGCATTACATTTTAAATATGCGAAGTGAAGAGACTGCAGCTGCAATAGCC 60  
 DB 717 GGCATTACATTTTAAATATGCGAAGTGAAGAGACTGCAGCTGCAATAGNC 776  
 QY 61 TAGGCTGAATTTTGTGCAGTAATAA 89  
 DB 777 TAGGGGTGAATTTTGTGCGGTGAATAA 805

RESULT 4  
 X20248-08  
 Continuation (9 of 10) of X20248 from base 800001 (Borrelia burgdorferi polynucleotide  
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
 WP Fragment Name Begin End  
 WP X20248-00 1 110000  
 WP X20248-01 100001 110000  
 WP X20248-02 200001 310000  
 WP X20248-03 300001 410000  
 WP X20248-04 400001 510000  
 WP X20248-05 500001 610000  
 WP X20248-06 600001 710000  
 WP X20248-07 700001 810000  
 WP X20248-08 800001 910000  
 WP X20248-09 900001 910715

Query Match 32.1%; Score 34; DB 1; Length 110000;  
 Best Local Similarity 61.1%; Pred. No. 0.13;  
 Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 5 TTCACATTTTAAATATGCGAAGTGAAGAGACTGCAGCTGCAATAGCC 64  
 DB 22717 TACAAAATATATCTTATCTGTTTTCATATGAGGGTTTGCAATTTCAATGCGCATG 22776  
 QY 65 GCTGAATTTTGTGCAGTAATAAATAA 94  
 DB 22777 TTGGAAATTTTTCAGACAGTAATAAGAA 22806

RESULT 5  
 ID 077884/c  
 AC 077884 standard; cDNA: 2520 BP.  
 DT 06-JUL-1995 (first entry)  
 DE Neutral thread protein Ad16c cDNA.  
 KM Neutral thread protein Ad16c; Alzheimer's; neuroectodermal tumours;  
 OS malignant astrocytomas; glioblastomas; ss.  
 FH Rattus rattus.  
 PN W09423756-A.  
 PD 27-OCT-1994.  
 PF 20-APR-1994; U04321.  
 PR 20-APR-1993; US-050559.  
 PA (GEHO) GEN HOSPITAL CORP.  
 PI De LA MONTE SM, Wands JR;  
 DR WPI: 94-341497/42.  
 DR Detection of neutral thread proteins - to detect sporadic and familial Alzheimer's disease, neuroectodermal tumours, malignant astrocytomas and glioblastomas (Eng).  
 PT Example 4; Fig 16S; 158pp; English.  
 CC 077884 is the Ad16c neutral thread protein (NTP) cDNA. This sequence was used in the development of an antibody dependent method, for the detection of NTPs. This new method could be used to diagnose Alzheimer's disease (differentiating between sporadic and familial), neuroectodermal tumours, malignant astrocytomas and glioblastomas.  
 CC Sequence 2520 BP; 661 A; 577 C; 581 G; 701 T;

Query Match 31.3%; Score 33.2; DB 1; Length 2520;



Synthetic.

CC from an organism, amplifying nucleic acid molecules whose nucleotide

CC sequence is homologous to amplification primers derived from the  
 CC fragment of the S. pneumoniae genome to prime the amplification and  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating  
 CC fragments of the S. pneumoniae genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for S. pneumoniae.  
 SQ Sequence 6735 BP; 2122 A; 1141 C; 1415 G; 2057 T;

Query Match 26.4%; Score 28; DB 1; Length 6735;  
 Best Local Similarity 56.5%; Pred. No. 5.2;  
 Matches 52; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 8 AATTTTAAATTTATGTCAGATGAGAGACTGAGCTGCAATAGCGCTAGGCT 67  
 DB 5876 ACTTTTAAAGAGCTGCTGCTGAGTGAACAGATTAAGTGGCAATGCAATATGGCT 5935  
 OY 68 GATTTTGTGAGATTAATTAATTAATCAT 99  
 DB 5936 GAATGCTATTGAAATTTGAATTAATTAATTAAT 5967

## RESULT 9

X20248\_05  
 Continuation (6 of 10) of X20248 from base 500001 (Borrelia burgdorferi polynucleotide S  
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

WP	Fragment Name	Begin	End
WP	X20248_00	1	110000
WP	X20248_01	100001	210000
WP	X20248_02	200001	310000
WP	X20248_03	300001	410000
WP	X20248_04	400001	510000
WP	X20248_05	500001	610000
WP	X20248_06	600001	710000
WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

Query Match 26.4%; Score 28; DB 1; Length 110000;  
 Best Local Similarity 60.5%; Pred. No. 9.6;  
 Matches 46; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY 11 TTTTAAATTTATGTCAGATGAGAGACTGAGCTGCAATAGCGCTAGGCTGAA 70  
 DB 85856 TTTTAAATTTCTTTTGTGATTAGATTAAGAAATTAAGATTAAATTAAGAGATTTA 85915

OY 71 TTTTGTGAGATTAAT 86  
 DB 85916 TTTTACTTATTATT 85931

## RESULT 10

O04525  
 ID 004525 standard; DNA; 134525 BP.  
 AC 004525;  
 DT 01-OCT-1990 (first entry)  
 DE Total base sequence of rice plant chloroplast DNA.  
 KW Chloroplast; true grass; rice plant; ss.  
 OS Oryza sativa.  
 PN J02100682-A.  
 PF 12-APR-1990.  
 PF 07-OCT-1988; 251967.  
 PR 07-OCT-1988; JP-251967.  
 RA (MITR) Mitsui Toatsu Chem Inc.  
 DR WPI; 90-159709/21.  
 PT Chloroplast DNA of true grasses - used to produce various  
 PS DNA base sequences by decomposition of rice plant DNA.  
 PS Claim 1; Fig 1; 20pp; Japanese.  
 CC The sequence is that of the whole of rice chloroplast DNA.  
 SQ Sequence 134525 BP; 41249 A; 26129 C; 26331 G; 40816 T;

Query Match 26.0%; Score 27.6; DB 1; Length 134525;  
 Best Local Similarity 55.1%; Pred. No. 13;  
 Matches 54; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 9 CATTTTAAATTTATGTCAGATGAGAGACTGAGCTGCAATAGCGCTAGGCTG 68  
 DB 44132 CCTATTTTCTATTGATGATGAGGATACCTTATGCAACAGAAATTTAGGCTTC 44191  
 OY 69 AATTTTGTGAGATTAATTAATTAATCATTCCT 106  
 DB 44192 CTTTATTGCAATTAAGAGAAATTCCTTCCTT 44229

## RESULT 11

N81716  
 ID N81716 standard; cDNA; 4436 BP.  
 AC N81716;  
 DT 10-MAR-1993 (revised)  
 DT 23-OCT-1990 (first entry)  
 DE Human growth hormone receptor clone.  
 KW Growth hormone receptor; gigantism; acromegaly; ss.  
 OS Homo sapiens.  
 FH Key location/Qualifiers  
 FT signal\_peptide 1..98  
 FT mat\_peptide 99..  
 FT poly\_a\_signal 4342..4347  
 FT tag\_c

PD W08809818-A.  
 PF 15-DEC-1988.  
 PF 10-JUN-1988; U02006.  
 PR 12-JUN-1987; US-062542.  
 PA (GETH) Genentech Inc.  
 PI Hammonds RG, Leung DW, Spencer SA, Wood WI;  
 DR WPI; 88-368632/51.  
 DR P-PSDB; P81326.  
 PT New pure growth hormone receptor and binding protein - for treating  
 PT growth hormone abnormalities, and new encoding DNA sequences.  
 PS Disclosure; 1pp; English.  
 CC The clone was isolated from an adult liver cDNA lambda gt10 library.  
 CC The DNA can be inserted into an expression vector for prodn. of  
 CC recombinant GHR. This is used to treat GH-related disorders such as  
 CC gigantism and acromegaly. The DNA can be used to screen children  
 CC for GHR defects.  
 CC See also N81717-9.  
 SQ Sequence 4436 BP; 1414 A; 866 C; 872 G; 1284 T;

Query Match 26.0%; Score 27.6; DB 1; Length 4436;  
 Best Local Similarity 63.6%; Pred. No. 6.3;  
 Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 11 TTTTAAATTTATGTCAGATGAGAGACTGAGCTGCAATAGCGCTAGGCTGAA 70  
 DB 2177 TTTTAAATTTGTGAGAGCTTTAGTAGCAGATGATGCTTAAATATGTGGGTAA 2236

OY 71 TTTTGTG 76  
 DB 2237 TTTTGTG 2242

## RESULT 12

V57245/C  
 ID V57245 standard; DNA; 435 BP.  
 AC V57245;  
 DT 10-NOV-1998 (first entry)  
 DE Clone #1 from mutant multiresistant L. lactis arhc gene.  
 KW Stress resistance; acid; temperature; oxidative stress; mutant; IS91;  
 KW wild type; transport; electron; amino acid; peptide; DNA repair; milk;

KM RNA stability; guanosine phosphate biosynthesis; insertion sequence; ss;  
 KM deletion; recombination; mobile genetic element; lactic acid bacteria;  
 KM Lactococcus; Lactobacillus; Streptococcus; fermentation; yogurt; cheese.  
 OS MO9810102-A2.  
 PD 12-MAR-1998.  
 PF 05-SEP-1997; F01566.  
 PR 23-SEP-1996; FR-011555.  
 PR 06-SEP-1996; FR-010926.  
 PA (INRG) INST NAT RECH AGRONOMIQUE.  
 PI Duval PLH, Gruss AD, Maguin EGI, Rallu FC;  
 DR WPI: 98-193644/17.  
 PR New bacteria with increased resistance to stress - have mutation(s)  
 PT in specific genes, particularly lactic acid bacteria with increased  
 PT tolerance of acid, heat or oxidation, useful for fermentation, food  
 PT preservation etc.  
 PS Disclosure: Page 41; 66pp; French.  
 CC The invention relates to the generation of bacteria with better  
 CC resistance to stress, e.g. better resistance to acid conditions,  
 CC temperature and oxidative stress, than the wild type by introducing  
 CC into a gene at least 1 mutation that alters the normal activity of it,  
 CC or of some other gene in the same transcription unit. The gene is  
 CC especially involved in transport of electrons, amino acids, oligopeptides  
 CC or phosphate, in DNA repair, stability of RNA, biosynthesis of guanosine  
 CC phosphates or are new genes. Mutation of the gene is: (a) by  
 CC insertion/deletion of a DNA sequence by recombination (preferably using a  
 CC mobile genetic element, particularly in plasmid pgh:ISS1); (b) treatment  
 CC with mutagen or (c) by spontaneous mutation. The bacteria are  
 CC particularly lactic acid bacteria, specifically Lactococcus,  
 CC Lactobacillus or Streptococcus, and are useful for fermentation (of milk  
 CC or dairy products such as yogurt or cheese), to preserve foods and for  
 CC probiotic purposes. This is an example of a fragment of a mutated arnc  
 CC gene generated by the method of the invention.  
 CC Sequence 435 BP; 118 A; 76 C; 65 G; 176 T;

Query Match 26.0%; Score 27.6; DB 1; Length 435;

Best Local Similarity 58.5%; Pred. No. 3.8;

Matches 48; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 13 TTAATAATTTATGAGAGTAGAGAGACTGCTCATATAGCCTAGGCTGATTT 72

DB 357 TTAATAATTTATGAGAGTAGAGAGACTGCTCATATAGCCTAGGCTGATTT 298

OY 73 TTTGTACATTAATAATAATAA 94

DB 297 TTCTCAAAAAAATGACACAGA 276

RESULT 13

V20912 20912 standard; cDNA; 1123 BP.

AC V20912.

DE 15-JUL-1998 (first entry)

DE Human DBI/ACBP -like protein (DBIH) cDNA

KM Diazepam binding inhibitor/acyl-CoA binding protein; DBI/ACBP; DBIH;

KM paranganglia; catecholamine; ss.

OS Homo sapiens.

FT Key Location/Qualifiers

FT CDS 121..558

FT /\*tag= a

FT /product= "Human DBI/ACBP-like protein"

FN MO9806847-A1.

PD 19-FEB-1998.

PD 14-AUG-1997; U14379.

PR 16-AUG-1996; US-700626.

PA (INCY- ) INCYTE PHARM INC.

PI Au-Young J, Goli SK, Hillman JL;

DR P-PSDB; W52811.

PT Human diazepam binding inhibitor/acyl-coenzyme A binding protein and

PT related nucleic acid - vectors, transformed cells and antibodies, and

PT useful for delivering water-insoluble drugs, and for diagnosis and

PT treatment of para-ganglial abnormalities

PS Claim 3; Fig 1A-1C; 56pp; English.

CC The present sequence represents a cDNA encoding the human diazepam

CC inhibitor/acyl-CoA binding protein (DBI/ACBP) -like protein (DBIH).

CC DBIH cDNA was isolated from a human paranganglia tissue library

CC (PGANOT01). DBIH cDNA can have several uses which may include

CC diagnosis of paranganglia associated disease, in gene therapy or as a

CC probe for mapping the genomic sequence. DBIH protein may be used for

CC diagnosis and treatment of conditions involving abnormal function of

CC the paranganglia (including paranganglioma where it may regulate

CC biosynthesis and metabolism of catecholamines), for generating specific

CC antibodies, as a soluble carrier for normally insoluble drugs or for

CC drug screening.

CC Sequence 1123 BP; 352 A; 246 C; 293 G; 232 T;

Query Match 25.8%; Score 27.4; DB 1; Length 1123;

Best Local Similarity 65.6%; Pred. No. 5.4;

Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 33 GATAGAGAACTGACGCTGCTCATATAGCCTAGGCTGATTTTGTGAGATAATAATA 92

DB 973 GACTGGAATACTGACGCTGCTCATATAGCCTAGGCTGATTTTGTGAGATAATAATA 1032

OY 93 A 93

DB 1033 A 1033

RESULT 14

V65252/C 20912 standard; DNA; 6217 BP.

AC V65252.

DE 24-DEC-1998 (first entry)

DE DNA encoding S. pneumoniae Arp binding protein.

KM Streptococcus pneumoniae protein; recombinant; gene expression;

KM DNA chip; virulence; antibody; infection; detection; treatment; ss.

OS Streptococcus pneumoniae.

PD MO9826072-A1.

PD 18-JUN-1998.

PR 09-DEC-1997; U22578.

PR 13-DEC-1996; US-036281.

PA (ELIL) LILLY & CO ELI.

PI Balitz RH, Burgelt SG, Denoff BS, Hoskins JA, Jaskunas SR,

PI Mills BJ, Norris FH, Peery RB, Rokey PR, Rostock PR,

PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,

PI Young Bellido MT;

DR WPI: 98-348529/30.

DR P-PSDB; W80179, W80180, W80181, W80182.

PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips

PT for evaluating gene expression, and identification of virulence

PT genes

PS Claim 1; Pages 108-111; 33pp; English.

CC This DNA encodes a S. pneumoniae filamentous temperature sensitive,

CC Arp binding protein and three other proteins of unknown functions.

CC The invention provides DNA sequences (V65201 to V65304) from the

CC Streptococcus pneumoniae genome and corresponding protein sequences

CC (W80605 to W80728). A recombinant host containing a vector comprising any

CC of the above nucleic acids can be used for the recombinant expression of

CC the protein sequences. The invention also provides a DNA chip having

CC arrayed on it at least 15 base pair fragment of any one or more of these

CC DNA sequences. The DNA chip can be used methods for evaluating gene

CC expression in S. pneumoniae and for identifying virulence genes in

CC S. pneumoniae. Antibodies that selectively bind to the above proteins or

CC peptide fragments can be used to treat S. pneumoniae infection. The

CC antibodies can also be used to detect S. pneumoniae cells.

Sequence 6217 BP; 1908 A; 1156 C; 1340 G; 1813 T;

Query Match 25.8%; Score 27.4; DB 1; Length 6217;

Best Local Similarity 62.3%; Pred. No. 7.9;

Matches 43; Conservative 0; Mismatches 26; Indels 0; Gaps 0;



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:29 ; Search time 3075.15 Seconds  
(Without alignments)  
67.993 Million cell updates/sec

Title: US-09-049-696-17

Perfect score: 106  
Sequence: 1 GGCATTGCATTTTAAAAAT.....AAAATAATCATTCATCCTT 106

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST.\*  
1: em\_est1.\*  
2: em\_est2.\*  
3: em\_est3.\*  
4: em\_est4.\*  
5: em\_est5.\*  
6: em\_est6.\*  
7: em\_est7.\*  
8: em\_est8.\*  
9: em\_est9.\*  
10: em\_est10.\*  
11: em\_est11.\*  
12: em\_est12.\*  
13: em\_est13.\*  
14: em\_est14.\*  
15: em\_est15.\*  
16: em\_est16.\*  
17: em\_est17.\*  
18: em\_est18.\*  
19: em\_est19.\*  
20: em\_est20.\*  
21: em\_est21.\*  
22: em\_est22.\*  
23: em\_est23.\*  
24: em\_est24.\*  
25: em\_est25.\*  
26: em\_est26.\*  
27: em\_est27.\*  
28: em\_est28.\*  
29: em\_est29.\*  
30: em\_est30.\*  
31: em\_est31.\*  
32: em\_est32.\*  
33: em\_est33.\*  
34: em\_est34.\*  
35: em\_est35.\*  
36: em\_est36.\*  
37: em\_est37.\*  
38: em\_est38.\*  
39: em\_est39.\*  
40: em\_est40.\*  
41: em\_est41.\*  
42: em\_est42.\*  
43: em\_est43.\*  
44: em\_est44.\*  
45: em\_est45.\*  
46: em\_est46.\*  
47: em\_est47.\*  
48: em\_est48.\*  
49: em\_est49.\*  
50: em\_est50.\*  
51: em\_est51.\*  
52: em\_est52.\*  
53: em\_est53.\*

54: em\_est22.\*  
55: em\_est23.\*  
56: em\_est24.\*  
57: em\_est25.\*  
58: em\_est26.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	106	100.0	490	34	AA508854	AA508854 n122n10.s
C 2	106	100.0	716	49	AI660234	AI660234 w68902.x
C 3	106	100.0	734	49	AI660957	AI660957 w120008.x
C 4	106	100.0	392	50	AI687981	AI687981 wa77f06.x
C 5	104.4	98.5	161	48	AI582072	AI582072 ar96a11.x
C 6	103	97.2	285	20	D25727	D25727 HONGSO4094
C 7	94.8	89.4	501	51	AI721275	AI721275 as82h08.x
C 8	92.8	87.5	335	51	AI721121	AI721121 as73d08.x
C 9	85.4	80.6	378	31	AA296955	AA296955 EST112726
C 10	47.4	44.7	431	30	AA382884	AA382884 my34g09.r
C 11	47.4	44.7	418	35	AA596289	AA596289 vo26a08.r
C 12	47.4	44.7	552	37	AA688953	AA688953 vo04h07.r
C 13	47.4	44.7	421	37	AA691866	AA691866 vs12g06.r
C 14	47.4	44.7	386	37	AA692521	AA692521 vt59g08.r
C 15	47.4	44.7	482	37	AA711228	AA711228 vt70c11.r
C 16	47.4	44.7	488	37	AA733415	AA733415 vt75b06.r
C 17	47.4	44.7	388	37	AA734161	AA734161 vs19g06.r
C 18	47.4	44.7	463	39	AA839323	AA839323 vo40f07.r
C 19	37.2	35.1	629	34	AA518112	AA518112 v123c06.r
C 20	35.6	33.6	452	39	AA822200	AA822200 vp36d05.r
C 21	33.8	31.9	507	47	AI492398	AI492398 t127e08.x
C 22	32.4	30.6	482	36	AA634176	AA634176 ac75b07.s
C 23	31	29.2	502	39	AA813859	AA813859 ny14f05.s
C 24	30.6	28.9	409	23	H40625	H40625 yp50d03.s1
C 25	30.2	28.5	507	48	AU051820	AU051820 AU051820
C 26	30.2	28.5	305	49	AU053915	AU053915 AU053915
C 27	30.2	28.5	340	49	AU054084	AU054084 AU054084
C 28	30.2	28.5	263	49	AU054102	AU054102 AU054102
C 29	30	28.3	341	36	AA648424	AA648424 ns22c07.s
C 30	30	28.3	445	40	AA393238	AA393238 op75h07.s
C 31	29.4	27.7	433	27	AA017245	AA017245 ze52a07.s
C 32	29.4	27.7	414	27	AA017254	AA017254 ze52b08.s
C 33	29.4	27.7	426	28	AA088275	AA088275 z182g01.s
C 34	29.4	27.7	378	38	AA774038	AA774038 ac35g09.s
C 35	29.4	27.7	489	40	AA17783	AA17783 on82b02.s
C 36	29.4	27.7	441	43	AI217678	AI217678 qh26b10.x
C 37	29.2	27.5	328	38	AA783716	AA783716 c7c1a1.f
C 38	29.2	27.5	305	38	AA786664	AA786664 j9h09a1.f
C 39	29.2	27.5	351	43	AI210635	AI210635 j9h09a1.f
C 40	29.2	27.5	395	43	AI212672	AI212672 x9d10a1.f
C 41	29.2	27.5	423	43	AI212673	AI212673 x9d10a1.f
C 42	29.2	27.5	346	45	AI327611	AI327611 d0h02a1.f
C 43	29	27.4	646	37	AA691335	AA691335 vs14h06.r
C 44	29	27.4	550	49	AI633145	AI633145 t219f12.x
C 45	28.8	27.2	464	35	C24771	C24771 C24771 Rice

#### ALIGNMENTS

RESULT 1  
LOCUS AA508854 490 bp mRNA  
DEFINITION n122n10.s1 NCI\_CGAP\_C04 Homo sapiens CDNA clone IMAGE:968803 3',  
ACCSSION AA508854  
NID G2246357  
VERSION AA508854.1 GI:2246357

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
JOURNAL NCI-CGAP National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
COMMENT Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1397407.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.dio.llnl.gov/bdrip/image/image.html  
Insert Length: 856 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 404.  
Location/Qualifiers  
1. 490  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="968803"  
/clone\_lib="NCI-CGAP\_C04"  
/sex="pooled"  
/tissue\_type="colon"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from pooled colon  
tumor tissue, and was then primed with a Not I - oligo(dT)  
primer. Double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
This library is not normalized. Library constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 155 a 82 c 89 g 164 t  
ORIGIN

Query Match 100.0%; Score 106; DB 34; Length 490;  
Best Local Similarity 100.0%; Pred. No. 7.2e-20;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCATTACATTTTAAATATGATGAGTAGAGAGAACTGCAATGACC 60  
|||||  
Db 317 GGCATTACATTTTAAATATGATGAGTAGAGAGAACTGCAATGACC 258  
|||||

QY 61 TAGGGCTGAATTTTGTGAGATAAATAAATCAATCCTT 106  
|||||  
Db 257 TAGGGCTGAATTTTGTGAGATAAATAAATCAATCCTT 212  
|||||

RESULT 2  
A1660234 716 bp mRNA EST 10-MAY-1999  
LOCUS A1660234  
DEFINITION we8902.x1 Soares.Dieckgraefe\_colon\_NHCD Homo sapiens cDNA clone  
IMAGE:2346290.3 similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA  
sequence.  
ACCESSION A1660234  
NID 94763804  
VERSION A1660234.1 GI:4763804  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
JOURNAL NCI-CGAP National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
COMMENT Unpublished (1997)  
On Mar 10, 1998 this sequence version replaced gi:2949219.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40up from G160  
High quality sequence stop: 387.  
Location/Qualifiers  
1. 716  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="2346290"  
/clone\_lib="Soares\_Dieckgraefe\_colon\_NHCD"  
/tissue\_type="colonic mucosa from 3 patients with Crohn's  
disease"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5',  
GTTCACCAATCTGAGTAGGAGCGCCGCCGCTTTTCTTTTCTTTT 3'],  
TGTACCAATCTGAGTAGGAGCGCCGCCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieckbeim.wustl.edu); colonic mucosa represents a range of  
disease involvement from moderate to severe Crohn's  
disease; samples include both perforating (fistulas) and  
non-perforating samples. Library constructed by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 220 a 123 c 140 g 233 t  
ORIGIN

Query Match 100.0%; Score 106; DB 49; Length 716;  
Best Local Similarity 100.0%; Pred. No. 5.4e-20;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCATTACATTTTAAATATGATGAGTAGAGAGAACTGCAATGACC 60  
|||||  
Db 314 GGCATTACATTTTAAATATGATGAGTAGAGAGAACTGCAATGACC 255  
|||||

QY 61 TAGGGCTGAATTTTGTGAGATAAATAAATCAATCCTT 106  
|||||  
Db 254 TAGGGCTGAATTTTGTGAGATAAATAAATCAATCCTT 209  
|||||

RESULT 3  
A1660957 734 bp mRNA EST 10-MAY-1999  
LOCUS A1660957  
DEFINITION w120008.x1 Soares.Dieckgraefe\_colon\_NHCD Homo sapiens cDNA clone  
IMAGE:2351151.3 similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA  
sequence.  
ACCESSION A1660957  
NID 94764540  
VERSION A1660957.1 GI:4764540  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index  
Unpublished (1997)  
On Mar 20, 1998 this sequence version replaced gi:2980033.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 477.

## FEATURES

SOURCE

1.734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="X"  
/clone="IMAGE:2351151"  
/clone.lib="Soares.Dieckgraefe.colon.NHUC"  
/tissue\_type="colonic mucosa from 5 ulcerative colitis  
patients"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pRT3D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCATCTGAGTGGAGCGCCGCTACTCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pRT3D vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieck@im.wustl.edu); colonic mucosa represents a range of  
disease involvement from mild cryptitis to severe  
ulceration, fibrosis, and degeneration. Library  
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 200 a 143 c 159 g 228 t 4 others  
ORIGIN

Query Match 100.0%; Score 106; DB 49; Length 734;  
Best Local Similarity 100.0%; Pred. No. 6.4e-20;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCATTCACATTTTAAATATGAGTGGATAGGAACTGCAGCTGCAATAGCC 60  
|||||  
Db 110 GGCATTCACATTTTAAATATGAGTGGATAGGAACTGCAGCTGCAATAGCC 51  
QY 61 TAGGCTGAATTTTGTGAGATTAATAATATCATTCATCCTT 106  
|||||  
Db 50 TAGGCTGAATTTTGTGAGATTAATAATATCATTCATCCTT 5

RESULT 4  
A1687981 392 bp mRNA EST 27-MAY-1999  
LOCUS A1687981  
DEFINITION wa77f06.x1 Soares\_NFL\_T.GBC.S1 Homo sapiens cDNA clone  
IMAGE:3302211 3' similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA  
sequence.  
ACCESSION A1687981  
NID 94899275  
VERSION A1687981.1 GI:4899275  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 392)  
NCT-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2947973.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco.

## FEATURES

SOURCE

1.392  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2302211"  
/clone.lib="Soares\_NFL\_T.GBC.S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pRT3D-Pac (Pharmacia) with  
a modified polylinker; Site.1: Not I; Site.2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHF, and B-cell  
NCI-CGAP GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo.

BASE COUNT 125 a 69 c 67 g 131 t  
ORIGIN

Query Match 100.0%; Score 106; DB 50; Length 392;  
Best Local Similarity 100.0%; Pred. No. 7.7e-20;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCATTCACATTTTAAATATGAGTGGATAGGAACTGCAGCTGCAATAGCC 60  
|||||  
Db 281 GGCATTCACATTTTAAATATGAGTGGATAGGAACTGCAGCTGCAATAGCC 222  
QY 61 TAGGCTGAATTTTGTGAGATTAATAATATCATTCATCCTT 106  
|||||  
Db 221 TAGGCTGAATTTTGTGAGATTAATAATATCATTCATCCTT 176

RESULT 5  
A1582072 161 bp mRNA EST 06-APR-1999  
LOCUS A1582072  
DEFINITION ar96a11.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
IMAGE:2173436 3', mRNA sequence.  
ACCESSION A1582072  
NID 94567969  
VERSION A1582072.1 GI:4567969  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 161)  
Hiller,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,  
Kritman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
TITLE Unpublished (1997)  
JOURNAL On Mar 19, 1997 this sequence version replaced gi:1900677.  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 115.





Oy	61	TAGGGCTGATTTTGTGCAGATAAATAATTAATTCATTACCCCTT	106
Dd	63	TAGGGCTACACTTTTGTCTCTATTATACAAAATCATTACCTT	18
RESULT	8		
LOCUS	A1721121/c		
DEFINITION	A1721121	335 bp	mRNA
	ab36008.x1	Barnstead colon HPLRB7 Homo sapiens CDNA clone	EST
	IMAGE:23343351	3' similar to TR:088826 088826 GDB-5 PROTEIN.	; mRNA
sequence.			
ACCESSION	A1721121		
NID	95038377		
VERSION	A1721121.1	GI:5038377	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 335)		
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubnue,T., Geisel,G., Jost,S.,		
	Krizman,D., Kucada,T., Lacy,M., Le,N., Lemon,G., Matra,M.,		
	Martin,J., Moore,B., Schellenberg,K., Stepec,M., Tan,F.,		
	Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.		
TITLE	WashU-NCI human EST Project		
JOURNAL	Unpublished (1997)		
COMMENT	On Jun 22, 1998 this sequence version replaced gi:3246762.		
	Contact: Wilson RK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LNL ; contact the		
	IMAGE Consortium ( <a href="http://infoimage.llnl.gov">infoimage.llnl.gov</a> ) for further information.		
	Possible reversed clone: similarity on wrong strand		
	Seq primer: -40UP from Gibco		
	High quality sequence stop: 312.		
FEATURES	Location/Qualifiers		
SOURCE	1..335		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_1=IMAGE:23343351"		
	/clone_1ib="Barnstead colon HPLRB7"		
	/sex="Male"		
	/dev_stage="adult, age 25"		
	/lab_host="DH10B (phage resistant)"		
	/note="Organ: colon; Vector: pRTD-Pac (Pharmacia) with a		
	modified polylinker; Site:1: EcoRI; Site:2: NotI; 1st		
	strand cDNA was primed with a Not I - oligo(dN) primer [5'		
	TGTTACCATCTGAAGTGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTT		
	3']; double-stranded cDNA was ligated to Eco RI adaptors		
	[5' ATTCACTAGTAAT 3' and 5' ATTAGTAGG 3'], digested		
	with Not I and cloned into the Not I and Eco RI sites of		
	the modified pTV3 vector. Library constructed by Bob		
	Barnstead."		
BASE COUNT	98 a	57 c	71 g
ORIGIN			109 t
Query Match	87.5%	Score 92.8;	DB 51; Length 335;
Best Local Similarity	93.3%;	Pred. No. 3.2e-16;	
Matches	97; Conservative	0; Mismatches	7; Indels
		Gaps	0;
Oy	1	GGCATTACATTTAAAAATATGTGGAAGTGATAGGAACATGACGTGCAATAGCC	60
Dd	104	GGCATTACATTTAAAAATATGTGGAAGTGATAGGAACATGACGTGCAATAGCC	45
Oy	61	TAGGGCTGATTTTGTCTGATTAATAAATAATTCATTACCC	104
Dd	44	TAGGAAAAATTCCTCATGATAAATAAATAATTCATTACCC	1

RESULT	9
LOCUS	AA296955 378 bp mRNA EST 18-Apr-1997
DEFINITION	EST111726 Colon 1 Homo sapiens CDNA 5' end, mRNA sequence.
ACCESSION	AA296955
NID	91949515
VERSION	AA296955.1 GI:1949515
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 378) Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodde,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moren-Palmanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shibley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Benharik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dinke,D., Feng,D.F., Ferlie,A., Fischer,C., Hastings,G.A., He,M.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J.J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL	Nature 377 (6547 suppl), 3-174 (1995)
MEDLINE	96026280
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1288227. Other ESTs: THC167738 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlav@tigr.org For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index ( <a href="http://www.tigr.org/tcdb/hgi/hgi.html">http://www.tigr.org/tcdb/hgi/hgi.html</a> ) Seq primer: M13 Reverse.
FEATURES	Location/Qualifiers
SOURCE	1..378 /organism="Homo sapiens" /db_xref="ATCC (lnhost):127834" /db_xref="taxon:9606" /clone_11b="Colon 1" /dev_stage="adult" /note="Organ: colon; EcoRI: Site-2: XhoI"
BASE COUNT	112 a 79 c 64 g 119 t 4 others
ORIGIN	
Query Match	80.6%; Score 85.4; DB 31; Length 378;
Best Local Similarity	96.6%; Pred. No.3.2e-14;
Matches	86; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	1 GGCATTCACATTTAAAAATATGATGGAAGTGGATAGAGAACTGACAGTGTCAATAGCC 60       DB 265 GGCATTCACATTTAAAAATATGGAAGTGGATAGAGAACTGACAGTGTCAATAGNC 324 
QY	61 TAGGGCTGAATTTTGTGCAGTAATAAA 89       DB 325 TAGGGCTGAATTTTGTGCAGTAATAAA 353 

RESULT 10  
AA238284  
LOCUS  
DEFINITION  
AA238284  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AA238284 431 bp mRNA EST 03-MAR-1997  
m34909.r1 Barstead mouse pooled organs MPLB4 Mus musculus CDNA  
clone IMAGE:697792.5 similar to TR:61184066 G1184066  
CACIUM-ACTIVATED CHLORIDE CHANNEL. ; mRNA sequence.

AA238284  
91862488  
AA238284.1 GI:1862488  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 431)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Willson, R. and  
Waterston, R.  
The WashU-HHMI Mouse EST Project  
Unpublished (1996).  
On Sep 12, 1996 this sequence version replaced gi:1393580.

Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:431352  
Putative full length read  
vector to vector length is 439  
Seq primer: -28m13 rev2 ET from Amersham.

FEATURES  
Source  
Location/Qualifiers  
1..431  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:697792"  
/clone\_lib="Barstead mouse pooled organs MPLB4"  
/sex="mixed"  
/tissue\_type="pooled organs"  
/dev\_stage="7 day"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT733-Pac (Pharmacia) with  
a modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st  
strand cDNA was primed with a Not I...oligo(dT) primer [5  
TGTTCGATCTGGAAGTGGAGCGGCGCCGCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(GTGATTCGCGTACC), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT733 vector.  
Library constructed by Bob Barstead."

BASE COUNT  
ORIGIN  
120 a 99 c 98 g 114 t

Query Match 44.7%; Score 47.4; DB 30; Length 431;  
Best Local Similarity 70.8%; Pred. No. 0.00069;  
Matches 63; Conservative 0; Mismatches 26; Indels 0; Gaps 0

QY 1 GGCATTACATTTTAAAAATATGTGGAAGTGATAGAGACGACGCTGTCATATGCC 60  
||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 229 GGCATCCACGTCGTGAAGTAATAGTGAAGTGCTAGGGGAATGACGATGACACTAGGT 288  
||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 61 TAGGGCTGAATTTTCTCAGATTAATAAA 89  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db .289 TTGCACGTGAATTTTCAGGCACAGAAATCA 317

```

RESULT      11
LOCUS       AA596289
DEFINITION  AA596289 418 bp mRNA EST 19-SEP-1997
            vo26608.01 Barstead mouse irradiated colon MRLuR7 Mus musculus cdna
            clone IMAGE:105098 5' similar to TR:G1184066 G1184066
            CALCITRIOL-INDUCED CHLORIDE CHANNEL. ; mRNA sequence.
ACCESSION   AA596289
NID         92411724
VERSION     AA596289.1 GI:2411724
KEYWORDS    EST.
SOURCE      house mouse.
            Mus musculus
ORGNAMTSM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 418)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellensberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,
            Theisberg,B., Wille,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
            The WashU-HMNI Mouse EST Project
            Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1290647.

TITLE       JOURNAL
FEATURES
SOURCE      1. .418
            Location/Qualifiers

```

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1050998"
/clone_lib="Barstead mouse irradiated colon MPlRB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
15TGTTCGACATCTGAGTGGAGCGGCCCCCTTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (AATTCGATCCTTG), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library constructed by Bob Barstead."
BASE COUNT      111 a      96 c      102 g      109 t
ORIGIN
Query Match      44.7%; Score 47.4; DB 35; Length 418;
Best Local Similarity 70.8%; Pred. No. 0.0007;
Matches 63; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
1 GGCATTCACATTTTAAATAATATATGTGGAAGTGGATAGAGAGAACTGCACATCAATACC 60
|||||
|||||
|||||
|||||
|||||
228 GGCATCCACAGCGCGTGAACATATATGTGGAAGTGGCGTGGAGGAATGCAGTGCACCTAGGT 287
|||||
|||||
|||||
|||||
|||||
61 TAGGCGCTGAATTTTGTGCAGATTAATAA 89
|||||
|||||
|||||
|||||
|||||
288 TTGCACATGATTTTTCAGCGCAAGAATCAA 316
|||||
|||||
|||||
|||||
|||||

```





Mon Aug 9 13:26:07 1999

us-09-049-696-17.rst

Page 9

Job time: 13009 sec

---

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:41:12; Search time 650.92 Seconds

(without alignments)  
93.017 Million cell updates/sec

Title: US-09-049-696-16

Perfect score: 242

Sequence: 1 GTTATTCTCCACAGACTC.....ATGATTTCAGCTTCCTGT 242

Scoring table: OLIGO\_NWC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	164	67.8	255	1 T22483	Human gene signatu
2	63	26.0	878	1 T45884	Human colon specif
3	63	26.0	878	1 V16672	Polynucleotide seq
4	18	7.4	14286	1 X13099	Enterococcus faeca
5	17	7.0	18177	1 N90490	DNA of human retin
6	17	7.0	110000	1 V21209_02	Continuation (3 of
7	17	7.0	110000	1 V21209_14	Continuation (15 o
8	17	7.0	18303	1 X04502	Human retinoblasto
9	16	6.6	1153	1 N91687	Human pituitary CD
10	16	6.6	5095	1 Q05590	Insert sequence en
11	16	6.6	8247	1 N70996	Sequence encoding
12	16	6.6	174	1 O33628	Mycoplasma-specifi
13	16	6.6	10718	1 O51476	Downstream sequenc
14	16	6.6	194	1 Q77356	Human genome fragm
15	16	6.6	4474	1 Q79703	Cyclic guanosine m
16	16	6.6	1153	1 T01862	Human 7B2 pre-prot
17	16	6.6	3000	1 T01402	Canine herpesvirus
18	16	6.6	6042	1 T09018	Arabidopsis thalia
19	16	6.6	4747	1 T09019	Arabidopsis thalia
20	16	6.6	960	1 T67981	H. pylori cytoplas
21	16	6.6	438	1 T67775	H. pylori cytoplas
22	16	6.6	438	1 T77455	H. pylori cytoplas
23	16	6.6	10723	1 T49304	CDNA encoding poly
24	16	6.6	10723	1 T49303	CDNA sequence enco
25	16	6.6	10723	1 T49303	CDNA sequence enco
26	16	6.6	37	1 T76027	DEN-2 cloning/sequ
27	16	6.6	59	1 T76028	DEN-2 cloning/sequ
28	16	6.6	52	1 T76029	DEN-2 cloning/sequ
29	16	6.6	4426	1 V09064	Bovine CGMP-bindin
30	16	6.6	5432	1 V10236	Arabidopsis cinnam
31	16	6.6	110000	1 V21209_00	Mechanococcus jan
32	16	6.6	4746	1 V57455	Arabidopsis ethyle
33	16	6.6	6172	1 V57454	Arabidopsis ethyle
34	16	6.6	237326	1 V57903	Hereditary haemoch
35	16	6.6	1086	1 V59767	Human secreted pro
36	16	6.6	1256	1 V59618	S. aureus D-alanin
37	16	6.6	2948	1 V08568	Murine JNK-Interac
38	16	6.6	2832	1 V69289	Staphylococcus aur
39	16	6.6	11126	1 V75590	Staphylococcus aur
40	16	6.6	11126	1 V74482	Staphylococcus aur
41	16	6.6	32768	1 X13065	Enterococcus faeca
42	16	6.6	3903	1 X14480	H. pylori GHPD 109
43	16	6.6	2093	1 X20305	Borrelia burgdorfe

## ALIGNMENTS

C	44	16	6.6	648	1	X21123
	45	16	6.6	830	1	X40052
Polynucleotide seq						
Prostate cancer as						
RESULT 1						
RESULT	1					
ID	T22483					
AC	T22483:					
DT	22-AUG-1996 (first entry)					
DE	Human gene signature H0XG504094.					
KW	Human signature; messenger RNA; mRNA; relative abundance; frequency;					
KW	human; cloning; mapping; non-biased library; diagnosis; detection;					
KW	cell typing; abnormal cell function; ss.					
OS	Human sapiens.					
PN	W0951472-AL.					
PD	01-JUN-1995.					
PF	11-NOV-1994; J01916.					
PR	12-NOV-1993; JP-355504.					
PA	(MATS/) MATSUBARA K.					
PA	(OKUBO/) OKUBO K.					
PI	Matsubara K, Okubo K;					
DR	WPI: 95-206931/27.					
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.					
PT	for diagnosis of abnormal cell function, by preparing cDNA that					
PT	reflects relative abundance of corresp. mRNA in specific human					
PT	tissues					
PS	Claim 1: Page 1138-1139; 2245pp; Japanese.					
CC	A single-stranded DNA (or its complementary strand or the corresp.					
CC	double-stranded DNA) which comprises one of the 7837 "GS" sequences					
CC	given in 119001-126837 and which is able to hybridise to part of					
CC	human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)					
CC	sequences were obtained from 3'-directed cDNA libraries prepared					
CC	from various human tissues; synthesis of cDNA was initiated from the					
CC	3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-					
CC	untranslated sequence is unique to a particular mRNA species, almost					
CC	all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library					
CC	is constructed so as to reflect accurately the relative abundance of					
CC	different mRNAs in the particular tissue from which it was derived.					
CC	The appearance frequency of a given GS in a cDNA library can be					
CC	determined (esp. using primers and probes derived from the GS					
CC	sequences) as a means of diagnosing abnormal cell function or for					
CC	recognising different cell types.					
SO	Sequence 255 BP; 76 A; 52 C; 39 G; 80 T;					

Query Match 67.8%; Score 164; DB 1; Length 255;

Best Local Similarity 100.0%; Pred. No. 3.7e-71;

Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	8	CCTCAGACAGCCGCGAGACGACCTAGCTGATGAAAGCTGCGCTTCTCTAT	67
DB	49	CCTCAGACAGCCGCGAGACGACCTAGCTGATGAAAGCTGCGCTTCTCTAT	108
QY	68	ATTTCATATCAACAGACACCATTCCTGCAATTCACATTTTAAATATATGGAAGTGA	127
DB	109	ATTTCATATCAACAGACACCATTCCTGCAATTCACATTTTAAATATATGGAAGTGA	168
QY	128	GGAGAACTGACAGCTGCAATAGCCTAGGCGTGAATTTTGTGAC	171
DB	169	GGAGAACTGACAGCTGCAATAGCCTAGGCGTGAATTTTGTGAC	212
RESULT 2			
ID	T45884		
AC	T45884:		
DT	13-MAR-1997 (first entry)		
DE	Human colon specific gene CSG5 cDNA partial clone.		
KW	Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;		
KW	gene therapy; ss.		

OS Homo sapiens. Location/Qualifiers  
 FH Key 2..691  
 FT cds /tag- a  
 PN MO9639419-A1.  
 PD 12-DEC-1996.  
 PF 06-JUN-1995; U07289.  
 PR 06-JUN-1995; WO-U07289.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Yu G;  
 P1 WPI: 97-043054/04.  
 DR P-PSDB: W06548.  
 PT Human colon specific genes and their expression products - detection of which, in non-colon tissue samples, can be used as indication of colon cancer metastasis  
 PS Claim 1; Fig 5; 60pp; English.  
 CC 13 CDNA clones (745880-92), most of them partial clones, correspond to human colon specific genes, designated CSG1, CSG2, etc., that are primarily expressed in tissues derived from the colon. CSG7 and CSG10 show reduced expression in colon cancer cells as compared to that in normal cells; the remaining genes are overexpressed in colon cancer. The partial CDNA sequences can be used to isolate full-length clones and genomic clones including the complete gene. CSG nucleic acids can be used to produce CS6 polypeptides (see also W06545-53) in transformed host cells, as probes to detect disorders of the colon, partic. colon cancer and colon cancer metastasis, and in gene therapy.  
 CC Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;  
 SQ

Query Match  
 Best Local Similarity 100.0%; Score 63; DB 1; Length 878;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 CCTAATATTCATATCAACGACCATTCCTGGCATTACATTTAAATATGCGAAG 121  
 DB 687 CCTAATATTCATATCAACGACCATTCCTGGCATTACATTTAAATATGCGAAG 746

OY 122 TGG 124  
 DB 747 TGG 749

RESULT 3  
 VI6672 standard; CDNA: 878 BP.  
 AC VI6672;  
 DT 22-JUN-1998 (first entry)  
 DE Polynucleotide sequence of a colon-specific gene.  
 KW Colon-specific gene; probe; detection; expression; human; diagnostic assay; colon cancer; antibody; screening; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 2..685  
 FT /tag- a  
 FT /note- "no stop codon given"  
 PN US5733748-A.  
 PD 31-MAR-1998.  
 PF 06-JUN-1995; 469667.  
 PR 06-JUN-1995; US-469667.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen C, Yu G;  
 P1 WPI: 98-229823/20.  
 DR P-PSDB: W46879.  
 PT Colon-specific nucleic acids - useful as probes for detecting colon cancer metastases  
 PS Claim 15; Fig 5A-B; 51pp; English.  
 CC VI6668-81 represent polynucleotide sequences of partial or full-length CDNA clones of colon-specific genes. The polynucleotides can be used as probes to detect expression of the corresponding human genes, e.g. in diagnostic assays for detecting micrometastases of colon cancer.  
 CC Recombinant cells containing the polynucleotides can be used to produce the polypeptides, in order that antibodies can be raised and

CC used in further screening or diagnostics.  
 SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;  
 SQ

Query Match  
 Best Local Similarity 100.0%; Score 63; DB 1; Length 878;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 CCTAATATTCATATCAACGACCATTCCTGGCATTACATTTAAATATGCGAAG 121  
 DB 687 CCTAATATTCATATCAACGACCATTCCTGGCATTACATTTAAATATGCGAAG 746

OY 122 TGG 124  
 DB 747 TGG 749

RESULT 4  
 XI3099/c standard; DNA: 14286 BP.  
 ID XI3099/c  
 AC XI3099;  
 DT 19-MAR-1999 (first entry)  
 DE Enterococcus faecalis genome contig SEQ ID NO:162.  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection; vaccine; attenuation; computer readable medium; ds.  
 OS Enterococcus faecalis.  
 PN W09850555-A2.  
 PD 12-NOV-1998.  
 PF 04-MAY-1998; U08985.  
 PR 14-NOV-1997; US-066009.  
 PR 06-MAY-1997; US-044031.  
 PR 16-MAY-1997; US-046655.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Dillon PJ, Kunsch CA;  
 P1 WPI: 99-045171/04.  
 DR New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT used to develop products for the detection of Enterococcus and for use in vaccines for prevention or attenuation of Enterococcus infection.  
 PS Claim 1; Page 907-914; 2084pp; English.  
 CC A computer readable medium has been developed which has recorded on it 982 nucleotide sequences isolated from the Enterococcus faecalis genome. X12938 to X13919 represent these nucleotide sequences which are primary nucleotide sequences, also known as contigs. The computer-based system can identify fragments of the Enterococcus faecalis genome with commercial importance. The products can be used to detect the presence of Enterococcus faecalis in samples. They can also be used for diagnosing Enterococcal infection in an animal and monitoring progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of Enterococcus faecalis, or another related organism, in vivo or in vitro. In particular the polypeptides encoded by the Enterococcus faecalis nucleotide sequences can be used in vaccines to prevent or attenuate an Enterococcal infection.  
 CC Sequence 14286 BP; 4262 A; 2884 C; 2303 G; 4821 T;  
 SQ

Query Match  
 Best Local Similarity 100.0%; Score 18; DB 1; Length 14286;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TCAGATAAATAAATAA 185  
 DB 10683 TCAGATAAATAAATAA 10666

RESULT 5  
 N90490/c standard; DNA: 18177 BP.  
 ID N90490/c  
 AC N90490;  
 DT 30-OCT-1989 (first entry)  
 DE DNA of human retinoblastoma gene exons with flanking regions.  
 KW Human; retinoblastoma gene; flanking regions; CDNA;



KM screening; tumours; probes; exons.  
OS Homo sapiens (human).  
FH Key Location/Qualifiers  
FH Key Location/Qualifiers  
FT exon  
FT intron (2.9 kb) 533..669  
FT intron (2.9 kb) 1008/1009  
FT intron (2.9 kb) 1318..1444  
FT intron (3 kb) 1712/1713  
FT intron (3 kb) 1995..2110  
FT intron (1.6 kb) 2363/2364  
FT intron (1.6 kb) 2559..2678  
FT intron (2.1 kb) 2933/2934  
FT intron (2.1 kb) 3227..3265  
FT intron (1.4 kb) 3549/3550  
FT intron (1.4 kb) 3695..3762  
FT intron (1.5 kb) 3983/3984  
FT intron (1.5 kb) 4285..4395  
FT intron (1.8 kb) 4639/4640  
FT intron (1.8 kb) 4909..5051  
FT intron (1.8 kb) 5248/5249  
FT intron (1.8 kb) 5480..5561  
FT intron (1.9 kb) 5734/5735  
FT intron (1.9 kb) 6000..6109  
FT intron (1.0 kb) 6381/6382  
FT intron (1.0 kb) 6690..6768  
FT intron (3.2 kb) 7061/7062  
FT intron (3.2 kb) 7375..7462  
FT intron (3.1 kb) 7765/7766  
FT intron (3.1 kb) 7996..8112  
FT intron (1.8 kb) 8349/8350  
FT intron (1.8 kb) 8680..8736  
FT intron (1.8 kb) 9140..9171  
FT intron (1.8 kb) 9252..9328  
FT intron (1.0 kb) 9629/9630  
FT intron (1.0 kb) 9879..10075  
FT intron (70.0 kb) 10332/10323  
FT intron (70.0 kb) 10564..10682  
FT intron (3.0 kb) 10911/10912  
FT intron (3.0 kb) 11225..11370  
FT intron (2.8 kb) 11561/11562  
FT intron (2.8 kb) 11809..11955  
FT intron (4.0 kb) 12278/12279  
FT intron (4.0 kb) 12358..12642  
FT intron (0.2 kb) 12948/12949  
FT intron (0.2 kb) 13204..13317  
FT intron (0.2 kb) 13411..13574  
FT intron (7.4 kb) 13883/13884  
FT intron (7.4 kb) 14209..14239  
FT intron (2.8 kb) 14641/14642  
FT intron (2.8 kb) 14834..14977  
FT intron (2.8 kb) 15488..15538  
FT intron (1.7 kb) 15952/15953  
FT intron (1.7 kb) 16143..16213  
FT intron (1.7 kb) 18010  
FT polyA signal  
PM W08906703-A.  
PD 27-JUL-1989.  
PF 23-JAN-1989; U00293.  
PR 21-JAN-1988; US-146525.  
PA (DRXJ) Dryja T P.  
PI Dryja T P; Friend S; Yandell D W.  
DR WPI789-233856/32.  
DR P-PSDB; P90599.  
PT New human retino-blastoma gene and polypeptide(s)  
PT gene and treating these patients.  
PS Disclosure: fig 6: 71 pp: English.  
CC DNA encoding human retinoblastoma (RB) gene exons and flanking regions.  
CC The cDNA (see P90599 and M90490) can be used to screen individuals  
CC for the presence of the mutated RB gene. The RB polypeptide  
CC can prevent retinoblastoma formation, and corresp. antibodies  
CC can be used in tumour immunodiagnosis. Refer to patent for more  
CC sequence analysis details.  
SQ Sequence 18177 BP; 5765 A; 2997 C; 3172 G; 6202 T; 41 Others.

Query Match 7.0%; Score 17; DB 1; Length 18177;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 98 CACATTTTAAATATAT 114  
DB 6298 CACATTTTAAATATAT 6282  
RESULT 6  
V21209\_02  
Continuation (3 of 17) of V21209 from base 200001 (Methanococcus jannaschii circular  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209  
WP Fragment Name Begin End  
WP V21209\_00 1 110000  
WP V21209\_01 100001 210000  
WP V21209\_02 200001 310000  
WP V21209\_03 300001 410000  
WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000  
WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000  
WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976  
Query Match 7.0%; Score 17; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 169 CAGATTAATTAATTA 185  
DB 69638 CAGATTAATTAATTA 69654  
RESULT 7  
V21209\_14  
Continuation (15 of 17) of V21209 from base 1400001 (Methanococcus jannaschii circular  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209  
WP Fragment Name Begin End  
WP V21209\_00 1 110000  
WP V21209\_01 100001 210000  
WP V21209\_02 200001 310000  
WP V21209\_03 300001 410000  
WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000  
WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000  
WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976  
Query Match 7.0%; Score 17; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 172 ATAAATTAATTAATCA 188



```

FT      /tag- 2
FT      /number-12
FT      /note- "the intron is 3.1 kb, there is a break in
FT      the sequence given in the specification,
FT      between nucleotides 7765 and 7766"
FT      7996. .8112
FT      /tag- aa
FT      /number-13
FT      /note- "deposited as ATCC 97950, specifically
FT      claimed in Claim 32"
FT      8113. .8679
FT      /tag- ab
FT      /number-13
FT      /note- "the intron is 1.8 kb, there is a break in
FT      the sequence given in the specification,
FT      between nucleotides 8349 and 8350"
FT      8680. .8736
FT      /tag- ac
FT      /number-14
FT      /note- "deposited as ATCC 97950, specifically
FT      claimed in Claim 33"
FT      8737. .9139
FT      /tag- ad
FT      /number-14
FT      9140. .9174
FT      /tag- ae
FT      /number-15
FT      /note- "deposited as ATCC 97950, specifically
FT      claimed in Claim 34"
FT      9175. .9251
FT      /tag- ae
FT      /number-15
FT      9252. .9328
FT      /tag- af
FT      /number-16
FT      /note- "deposited as ATCC 97950, specifically
FT      claimed in Claim 35"
FT      9329. .9878
FT      /tag- ag
FT      /number-16
FT      /note- "the intron is 1.0 kb, there is a break in
FT      the sequence given in the specification,
FT      between nucleotides 9629 and 9630"
FT      9879. .10076
FT      /tag- ah
FT      /number-17
FT      /note- "deposited as ATCC 97950, specifically
FT      claimed in Claim 36"
FT      10077. .10563
FT      /tag- ai
FT      /number-17
FT      /note- "the intron is 70.0 kb, there is a break in
FT      the sequence given in the specification,
FT      between nucleotides 10322 and 10323"
FT      10564. .10682
FT      /tag- aj
FT      /number-18
FT      /note- "deposited as ATCC 97949, specifically
FT      claimed in Claim 37"
FT      10683. .11224
FT      intron

Query Match      7.0%; Score 17; DB 1; Length 18303;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      98 CACATTAAATAATAT 114
        |||||
Db      6362 CACATTAAATAATAT 6346

RESULT 9
N91687/c
ID      N91687 standard; cDNA; 1153 BP.

```

```

AC      N91687;
DE      12-MAY-1990 (first entry)
DE      Human pituitary cDNA encoding 7B2 preproteins
KW      7B2 preproteins; human pituitary; small cell lung tumour.
OS      Homo sapiens.
FH      Key      Location/Qualifiers
FT      cds      28. .663
FT      /tag- a
FT      /number-13
FT      /tag- b
FT      /number-13
FT      /tag- c
FT      /number-13
FT      EP-315254-A.
FT      10-MAY-1989.
FT      28-OCT-1988; 202395.
FT      30-OCT-1987; NL-002590.
FT      (UYKA-) Sticht Kathol Univ.
FT      Martens GJM;
FT      WPI; 89-139252/19.
FT      P-PSDB; P91077.
FT      Recombinant human 7B2 protein
FT      - used as marker for small cell lung- and endocrine
FT      tumours and for producing antibodies for immuno-analysis
FT      Figure 1; page 12; 15pp; English.
FT      It encodes human 7B2 preproteins with an 18-26 amino acid signal peptide
FT      sequence and a 185 amino acid mature 7B2 protein. 7B2 is a marker for
FT      CC classic small cell lung tumours which may be detd. using labelled RNA or
FT      CC DNA probes or the Abs.
FT      Sequence 1153 BP; 323 A; 241 C; 283 G; 306 T;

Oy      95 ATTCACATTTTAAAA 110
        |||||
Db      887 ATTCACATTTTAAAA 872

Query Match      6.6%; Score 16; DB 1; Length 1153;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
ID      005590 standard; cDNA; 5095 BP.
AC      005590;
DE      14-DEC-1990 (first entry)
DE      Insert sequence encoding signal peptide and first 1643 AAs of
DE      mature apob protein.
KW      Atherosclerosis; lipid binding peptide; LBP; phospholipid;
KW      lecithin cholesterol acyl transferase; L-CAT; apo AI;
KW      Chinese hamster ovaries.
OS      Homo sapiens.
FH      Key      Location/Qualifiers
FT      cds      85. .5095
FT      /tag- a
FT      /number-13
FT      /tag- b
FT      /number-13
FT      signal_peptide 85. .165
FT      US4943527-A.
FT      24-JUL-1990.
FT      27-FEB-1986; 834300.
FT      04-OCT-1985; US-784418.
FT      04-DEC-1985; US-804692.
FT      27-FEB-1986; US-834300.
FT      (CALB-) CALIFORNIA BIOTECHN.
FT      Procter AA, Vigne JL, Mallory JB, Talmadge KD, Kane JP;
FT      WPI; 90-246622/32.
FT      P-PSDB; R06341.
FT      Prodn. of purified lipid-binding peptide - capable of binding to
FT      FT phospholipid(s), useful in e.g. atherosclerosis.
FT      Disclosure; pp; English.
FT      Lipid-binding peptides in conjunction with phospholipids are
FT      CC effective in reversing cholesterol transport, and may enhance
FT      CC lecithin cholesterol acetyl transferase, lowering the rate of
FT      CC plaque formation and atherosclerosis.

```

Sequence 5095 BP: 1430 A; 1281 C; 1179 G; 1205 T;

Query Match 6.6%; Score 16; DB 1; Length 5095;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CTCACAGACTCCGCC 24

DB 3042 CTCACAGACTCCGCC 3057

# RESULT 11

N70996 N70996 standard; cDNA; 8247 BP.

AC N70996;

DT 17-APR-1991 (first entry)

DE Sequence encoding N-terminal apolipoprotein B (apob).

KW Lipid-binding peptide; apolipoprotein purification; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT signal\_peptide 87..167

FT /\*tag= a

FT mat\_peptide 168..8247

FT /\*tag= b

PN MO8702062-A.

PD 09-APR-1987.

PF 02-OCT-1985; US-784418.

PR 04-DEC-1985; US-804692.

PR 27-FEB-1986; US-834300.

PA (BIOT-) BIOTECHN RES PARTN.

PI Proteol AA, Vigne JL, Mallory JB, Talmadge KD;

DR WPI: 87-108703/15.

DR P-PEDB: P70647.

PT Purified lipid-binding peptide prodn. - by expression of

PT recombinant system and adding lipid to obtd. peptide mixt. to

PT form lipo-peptide complex

PS Disclosure: pp73-85; 123pp; English.

CC Apob is a lipid-binding peptide adapted to bind to phospholipids at

CC one or more amphipathic alpha-helical peptide regions. Apob gene is

CC pref. expressed in Chinese hamster ovary (CHO) cells and the gene is

CC introduced in an expression vector having a regulatable promoter

CC derived from the human metallothionein II gene.

CC Sequence 8247 BP; 2497 A; 1933 C; 1768 G; 2045 T;

QY Query Match 6.6%; Score 16; DB 1; Length 8247;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CTCACAGACTCCGCC 24

DB 3042 CTCACAGACTCCGCC 3057

# RESULT 12

Q13231 013231 standard; DNA; 890 BP.

AC 013231;

DT 29-OCT-1991 (first entry)

DE Mycoplasma-specific DNA.

KW Probe; primer; ss.

OS Mycoplasma.

PN J03161000-A.

PD 10-JUL-1991.

PF 16-NOV-1989; 296258.

PR 16-NOV-1989; JP-296258.

PR (TAKA-) TAKARA SHUZO KK.

PA WPI: 91-248474/34.

PT Mycoplasma detection - Involves detecting specified DNA sequence

PT of mycoplasma using high sensitivity.

PS Claim 1; Page 1; 8pp; Japanese.

CC The sequence is used to design primers and probes for the detection

CC of Mycoplasma with high sensitivity.

CC See also Q13232-Q13234.

Sequence 890 BP; 300 A; 131 C; 113 G; 346 T;

QY Query Match 6.6%; Score 16; DB 1; Length 890;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 TAAATTTCTAAAT 224

DB 329 TAAATTTCTAAAT 344

# RESULT 13

Q33628 033628 standard; DNA; 174 BP.

AC Q33628;

DT 02-FEB-1993 (first entry)

DE Downstream sequence of microsatellite from clone GBRKAS.

KW PCR; selection; primers; OPRIPRIM; breeding; cattle; parentage;

OS Bos taurus.

PN N09213102-A.

PD 06-AUG-1992.

PF 15-JAN-1991; US-642342.

PA (GENM-) GENMARK.

PI Georges M, Massey JM;

DR WPI: 92-284684/34.

PT Polymorphic bovine DNA markers - used in genetic identification,

PT gene mapping, and selective breeding

PS Table 7; Page 180; 51pp; English.

CC The sequence is that downstream of a bovine microsatellite sequence

CC obtd. by screening a library of bovine MBOI DNA fragments of between

CC 250 and 500 bp with an (AC)<sub>15</sub> and a (TC)<sub>15</sub> oligonucleotide probe.

CC One out of 50 clones cross-hybridised. Assuming independent

CC distribution of microsatellites and MBOI sites, the frequency of

CC (76)<sub>n</sub> >9 microsatellites in the bovine genome is estimated at >100,

CC 000. The sequence information for ca. 230 such bovine microsatellites

CC is summarised in the specification and indexed herein (see below).

CC The sequences upstream and downstream of the microsatellite sequence

CC were used to generate the required PCR primers for in vitro

CC amplification of the corresp. microsatellite (using the program

CC OPRIPRIM). The microsatellites may be used to identify individuals,

CC for parentage testing, and in the genetic mapping of economic trait

CC loci, or genes involved in the determination of economically important

CC traits esp. in cattle, to allow selective breeding.

CC See also Q33501-34437.

CC Sequence 174 BP; 77 A; 31 C; 19 G; 47 T;

QY Query Match 6.6%; Score 16; DB 1; Length 174;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 AGATTAATTAATAA 185

DB 129 AGATTAATTAATAA 144

# RESULT 14

Q51476/c 051476 standard; DNA; 10718 BP.

AC Q51476;

DT 16-MAY-1994 (first entry)

DE DEN1-S275/90 (ECACC V92042111).

KW Dengue haemorrhagic fever; DHF; dengue fever; DF; dengue shock syndrome;

KW DSS; DEN1 polyepitides; ss.

OS Dengue Fever Virus Type 1 strain S275/90.

FT key Location/Qualifiers

FT cds 81..10271

```

Query Match 16: Score 16; DB 1; Length 10718;
Best Local Similarity 100.0%; Pred.No. 53;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 75 TCACAGCAGCACCATCC 90
    |||||
Db 10705 TCACAGCAGCACCATCC 10690

RESULT 15
Q77356/c
ID Q77356 standard; DNA; 194 BP.
AC Q77356;
DT 22-NOV-1994 (first entry)
DE Human genome fragment (Preferred)

```

	Query Match	Score 16;	DB 1;	Length 194;
	Best Local Similarity	100.0%;	Pred. No. 53;	
	Matches 16; Conservative	0;	Mismatches	0; Indels
	Gaps	0;		
OY	95 ATTCACATTTRAAAAA	110		
DB	142 ATTCACATTTRAAAAA	127		

Search completed: August 5, 1999, 17:41:32  
Job time: 6327 sec

Mon Aug 9 13:26:53 1999

us-09-049-696-16\_1.rng

---

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:26 ; Search time 3075.15 Seconds  
(without alignments)  
155.229 Million cell updates/sec

Title: US-09-049-696-16

Perfect score: 242  
Sequence: 1 GTTATTTCCTCACAGACTC.....ATGTTATTTCAGCTTCCTCT 242

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:\*

1: em\_est1:\*\n2: em\_est2:\*\n3: em\_est3:\*\n4: em\_est4:\*\n5: em\_est5:\*\n6: em\_est6:\*\n7: em\_est7:\*\n8: em\_est8:\*\n9: em\_est9:\*\n10: em\_est10:\*\n11: em\_est11:\*\n12: em\_est12:\*\n13: em\_est13:\*\n14: em\_est14:\*\n15: em\_est15:\*\n16: em\_est16:\*\n17: em\_est17:\*\n18: em\_est18:\*\n19: em\_est19:\*\n20: em\_est20:\*\n21: em\_est21:\*\n22: em\_est22:\*\n23: em\_est23:\*\n24: em\_est24:\*\n25: em\_est25:\*\n26: em\_est26:\*\n27: em\_est27:\*\n28: em\_est28:\*\n29: em\_est29:\*\n30: em\_est30:\*\n31: em\_est31:\*\n32: em\_est32:\*\n33: em\_est33:\*\n34: em\_est34:\*\n35: em\_est35:\*\n36: em\_est36:\*\n37: em\_est37:\*\n38: em\_est38:\*\n39: em\_est39:\*\n40: em\_est40:\*\n41: em\_est41:\*\n42: em\_est42:\*\n43: em\_est43:\*\n44: em\_est44:\*\n45: em\_est45:\*\n46: em\_est46:\*\n47: em\_est47:\*\n48: em\_est48:\*\n49: em\_est49:\*\n50: em\_est50:\*\n51: em\_est51:\*\n52: em\_est52:\*\n53: em\_est53:\*

54: em\_est23:\*\n55: em\_est24:\*\n56: em\_est25:\*\n57: em\_est26:\*\n58: em\_est27:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	242	100.0	490	34	AA508854	AA508854 n122h10.s
C 2	242	100.0	716	49	AI660234	AI660234 w68g02.x
C 3	226.8	93.7	392	50	AI687981	AI687981 wa77f06.x
C 4	201.2	83.1	255	20	D25727	D25727 HUMS04094
C 5	199	82.2	734	49	AI660957	AI660957 w120d08.x
C 6	194.6	80.4	501	51	AI721275	AI721275 as82h08.x
C 7	183.8	76.0	335	51	AI721121	AI721121 as73c08.x
C 8	181	74.8	378	31	AA296955	AA296955 EST112726
C 9	155.8	64.4	161	48	AI582072	AI582072 ar96a11.x
C 10	92	38.0	431	30	AA238284	AA238284 my34g09.x
C 11	92	38.0	418	35	AA596289	AA596289 vo26a08.x
C 12	92	38.0	421	37	AA691586	AA691586 vs12g06.x
C 13	92	38.0	386	37	AA692521	AA692521 vt59g08.x
C 14	92	38.0	482	37	AA711228	AA711228 vt70c11.x
C 15	92	38.0	488	37	AA733415	AA733415 vt75b06.x
C 16	92	38.0	388	37	AA734161	AA734161 vs19g06.x
C 17	92	38.0	463	39	AA839323	AA839323 vo40f07.x
C 18	87	36.0	552	37	AA688953	AA688953 vs04h07.x
C 19	73	30.2	629	34	AA518112	AA518112 v123c06.x
C 20	65	26.9	452	39	AA822200	AA822200 vp36d05.x
C 21	64.8	26.8	646	37	AA691335	AA691335 vs14h06.x
C 22	49	20.2	297	49	AI641931	AI641931 vt50e03.x
C 23	40.6	16.8	550	49	AI633145	AI633145 tz19f12.x
C 24	36.2	15.0	193	25	W09682	W09682 ma03h06.r1
C 25	36.2	15.0	431	25	W12467	W12467 ma55d09.r1
C 26	36.2	15.0	223	26	W17683	W17683 mb77e02.r1
C 27	36.2	15.0	318	33	AA419977	AA419977 vf54f09.x
C 28	36.2	15.0	413	34	AA63138	AA63138 vt70h10.x
C 29	36.2	15.0	241	36	AA646688	AA646688 ve44h02.x
C 30	36.2	15.0	575	36	C78440	C78440 C78440 Mous
C 31	36.2	15.0	580	36	C78445	C78445 C78445 Mous
C 32	36.2	15.0	470	44	AI265130	AI265130 uk03e10.y
C 33	36.2	15.0	665	49	AV001311	AV001311 AV001311
C 34	36.2	15.0	424	49	AV020365	AV020365 AV020365
C 35	35.2	14.5	234	40	AA929413	AA929413 v253h11.x
C 36	34.8	14.4	502	33	AA428896	AA428896 zw51c03.x
C 37	34.8	14.4	527	39	AA866537	AA866537 TENU0109
C 38	34.6	14.3	247	20	T18166	T18166 0759c3 czap
C 39	34.6	14.3	193	40	C90663	C90663 C90663 Dict
C 40	34.6	14.3	237	40	C92265	C92265 C92265 Dict
C 41	34.6	14.3	236	44	AU038794	AU038794 AU038794
C 42	34.6	14.3	318	49	AV016126	AV016126 AV016126
C 43	34.6	14.3	322	49	AV022659	AV022659 AV022659
C 44	34.4	14.2	410	41	AI067563	AI067563 EST209245
C 45	34.2	14.1	499	51	AI733365	AI733365 op05f04.x

## ALIGNMENTS

RESULT 1\nLOCUS AA508854 490 bp mRNA\nDEFINITION n122h10.s1 NCI\_CGAP\_Co4 Homo sapiens cDNA clone IMAGE:968803 3',\nACCESSION AA508854\nVERSION g2246357\nAA508854.1 GI:2246357

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 490)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397407.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/INL at:  
[www.bio.lnl.gov/bbrp/image/image.html](http://www.bio.lnl.gov/bbrp/image/image.html)  
Insert Length: 856 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 404.  
Location/Qualifiers  
1. 490  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:968803"  
/clone\_lib="NCI-CGAP\_C04"  
/sex="pooled"  
/tissue\_type="colon"  
/lab\_host="DH10B"  
/note="Vector: pRT73D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from pooled colon  
tumor tissue, and was then primed with a Not I - oligo(dT)  
primer. Double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pRT73 vector.  
This library is not normalized. Library constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 155 a 82 c 89 g 164 t  
ORIGIN  
Query Match 100.0%; Score 242; DB 34; Length 490;  
Best Local Similarity 100.0%; Pred. No. 5.1e-45;  
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GTTATTCCTCCACAGACTCGCCAGAGACACTGACTGATGAAGAGTCTCTCTG 60  
DB 408 GTTATTCCTCCACAGACTCGCCAGAGACACTGACTGATGAAGAGTCTCTCTG 349  
OY 61 TCCATATTCATATCAACAGACACCACTTCCTGGCATTCACATTTTAAATATATGCGAA 120  
DB 348 TCCATATTCATATCAACAGACACCACTTCCTGGCATTCACATTTTAAATATATGCGAA 289  
OY 121 GTGATAGAGAACTGCAGCTGTCAATAGCCTAGGGCTGAATTTTGTGAGATTAATAA 180  
DB 288 GTGATAGAGAACTGCAGCTGTCAATAGCCTAGGGCTGAATTTTGTGAGATTAATAA 229  
OY 181 ATAAATCATTCATCTTTTGTATTAATAATTTTCTAAATATGATTTTGAAGCTTCT 240  
DB 228 ATAAATCATTCATCTTTTGTATTAATAATTTTCTAAATATGATTTTGAAGCTTCT 169  
OY 241 GT 242  
DB 168 GT 167

RESULT 2  
LOCUS AT660234/c  
DEFINITION we88902.x1 Soares Dieckgraefe.colon.NHCD Homo sapiens cDNA clone  
IMAGE:2346290 3 similar to TR:088826 088826 G08-5 PROTEIN.; mRNA  
sequence.  
ACCESSION AT660234  
NID 94763804  
VERSION AT660234.1 .GI:4763804  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 716)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949219.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through INL; contact the  
IMAGE Consortium ([infoimage.lnl.gov](http://infoimage.lnl.gov)) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 387.  
Location/Qualifiers  
1. 716  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2346290"  
/clone\_lib="Soares Dieckgraefe.colon.NHCD"  
/tissue\_type="colonic mucosa from 3 patients with Crohn's  
disease"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pRT73D-Pac (Pharmacia) with a  
modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer (5'  
TGTTACCAATTCGTAAGTGGAGCGCGCGCTTTTATTTT 3'),  
TGTTACCAATTCGTAAGTGGAGCGCGCGCTTTTATTTT 3'),  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pRT73 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieckgraef.wustl.edu); colonic mucosa represents a range of  
disease involvement from moderate to severe Crohn's  
disease; samples include both perforating (fistulas) and  
non-perforating samples. Library constructed by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 220 a 123 c 140 g 233 t  
ORIGIN  
Query Match 100.0%; Score 242; DB 49; Length 716;  
Best Local Similarity 100.0%; Pred. No. 4.7e-45;  
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GTTATTCCTCCACAGACTCGCCAGAGACACTGACTGATGAAGAGTCTCTCTG 60  
DB 405 GTTATTCCTCCACAGACTCGCCAGAGACACTGACTGATGAAGAGTCTCTCTG 346  
OY 61 TCCATATTCATATCAACAGACACCACTTCCTGGCATTCACATTTTAAATATATGCGAA 120  
DB 345 TCCATATTCATATCAACAGACACCACTTCCTGGCATTCACATTTTAAATATATGCGAA 286  
OY 121 GTGATAGAGAACTGCAGCTGTCAATAGCCTAGGGCTGAATTTTGTGAGATTAATAA 180  
DB 285 GTGATAGAGAACTGCAGCTGTCAATAGCCTAGGGCTGAATTTTGTGAGATTAATAA 226  
OY 181 ATAAATCATTCATCTTTTGTATTAATAATTTTCTAAATATGATTTTGAAGCTTCT 240



Db 225 ATAAATCATTCATCCCTTTTGGATTATATAAATTTCTAAATGTTAGACTTCCT 166

QY 241 GT 242  
11  
Db 165 GT 164

RESULT 3  
LOCUS A1687981 392 bp mRNA EST 27-MAY-1999  
DEFINITION wa77f06.x1 Soares\_NFL.T.GBC.S1 Homo sapiens CDNA clone  
IMAGE:2302211 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA  
sequence.  
ACCESSION A1687981  
NID 94899275  
VERSION A1687981.1 GI:4899275  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 392)  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
COMMENT Tumor Gene Index  
Unpublished (1997)  
On Mar 10, 1998 this sequence version replaced gi:2947973.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LML; contact the  
IMAGE Consortium (info@image.lml.gov) for further information.  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. 392  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2302211"  
/clone\_lib="Soares\_NFL.T.GBC.S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pTT3D-Pac (Pharmacia) with  
a modified polylinker. Site.1: Not I; Site.2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbHL19W, testis NHT, and B-cell  
NCI-CGAP.GCB1) were mixed, and ss circles were made in  
vitro. Following RAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
1.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 125 a 69 c 67 g 131 t

ORIGIN

Query Match 93.7%; Score 226.8; DB 50; Length 392;  
Best Local Similarity 99.1%; Pred. No. 1.3e-41;  
Matches 228; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTATTTCTCCACAGACTCCGCCAGAGACACTAGTCTGATGAACGTCCTCCTTG 60  
|||||  
Db 372 GTTATTTCTCCACAGACTCCGCCAGAGACACTAGTCTGATGAACGTCCTCCTTG 313  
|||||

QY 61 TCCATATTCATATCAACAGCACTCTCTGCTGATCAATTTAAATTTATGTGGAA 120  
|||||  
Db 312 TCCATATTCATATCAACAGCACTCTCTGCTGATCAATTTAAATTTATGTGGAA 253  
|||||

QY 121 GTGATAGGAGAACTGCTCAATAGCCTAGGCTGAATTTTGTGCAGATAAATAA 180  
|||||  
Db 252 GTGATAGGAGAACTGCTCAATAGCCTAGGCTGAATTTTGTGCAGATAAATAA 193  
|||||

QY 181 ATAAATCATTCATCCCTTTTGGATTATATAAATTTCTAAATGTTATTT 230  
|||||  
Db 192 ATAAATCATTCATCCCTTTTGGATTATATAAATTTCTAAATGTTATTT 143  
|||||

RESULT 4  
LOCUS D25727 255 bp mRNA EST 30-NOV-1995  
DEFINITION HUMGS04094 Human colon mucosa Homo sapiens CDNA clone cm2037 3',  
mRNA sequence.  
ACCESSION D25727  
NID 9500422  
VERSION D25727.1 GI:500422  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 255)  
TITLE Okubo,K., Yoshii,J., Yokouchi,H., Kameyama,M. and Matsubara,K.  
Global analysis of gene expression in colon mucosa: a large scale  
random CDNA sequencing analysis  
Unpublished (1994)

CONTACT: Okubo,K., Itoh,K., Yoshii,J., Yokouchi,H. and Matsubara,K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.  
Location/Qualifiers  
1. 255  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cm2037"  
/clone\_lib="Human colon mucosa"  
/note="Adult male, tissue type = colon mucosa"

BASE COUNT 76 a 52 c 39 g 80 t

ORIGIN

Query Match 83.1%; Score 201.2; DB 20; Length 255;  
Best Local Similarity 95.8%; Pred. No. 7.4e-36;  
Matches 203; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCTGATGAACGTCCTCCTGT 61  
|||||  
Db 43 TTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCTGATGAACGTCCTCCTGT 102  
|||||

QY 62 CCTAATTCATATCAACAGCACTCTCTGCTGATCAATTTAAATTTATGTGGAG 121  
|||||  
Db 103 CCTAATTCATATCAACAGCACTCTCTGCTGATCAATTTAAATTTATGTGGAG 162  
|||||

QY 122 TGGATAGGAGAACTGCTCAATAGCCTAGGCTGAATTTTGTGCAGATAAATAA 181  
|||||  
Db 163 TGGATAGGAGAACTGCTCAATAGCCTAGGCTGAATTTTGTGCAGATAAATAA 222  
|||||

QY 182 TAAATCATTCATCCCTTTTGGATTATATAA 213  
|||||  
Db 223 TAAATCATTCATCCCTTTTGGATTATATAA 254  
|||||

RESULT 5  
LOCUS A1660957 734 bp mRNA EST 10-MAY-1999  
DEFINITION wf20d08.x1 Soares\_Dieckgraefe.colon.NHUC Homo sapiens CDNA clone  
IMAGE:2351151 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA  
sequence.  
ACCESSION A1660957  
NID 94764540  
VERSION A1660957.1 GI:4764540  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 734)  
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980033.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Seq primer: -400P from Glibco  
High quality sequence stop: 477.  
Location/Qualifiers  
1..734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="x"  
/clone="IMAGE:2351151"  
/clone.lib="Soares Dieckgraefe.colon\_NHUC"  
/tissue.type="colonic mucosa from 5 ulcerative colitis  
patients"  
/lab.host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCTGAGTGGAGCGCGCCGCTGCTGATTTTAAATTAATGAGAA  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified p773 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieck@im.wustl.edu); colonic mucosa represents a range of  
disease involvement from mild cryptitis to severe  
ulceration, fibrosis and degeneration. Library  
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 200 a 143 c 159 g 228 t 4 others

Query Match 82.2%; Score 199; DB 49; Length 734;  
Best Local Similarity 100.0%; Pred. No. 1.8e-35;  
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTATTTCTCCACAGACCTCCGCGCAGACACCTAGTCTGTGTAAGAAAGCTGCTCTCTG 60  
|||||  
DB 201 GTTATTTCTCCACAGACCTCCGCGCAGACACCTAGTCTGTGTAAGAAAGCTGCTCTCTG 142  
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 TCCATAATTCATATACACAGACCAATTCCTGCGCATTCACATTTTAAATTAATGAGAA 120  
|||||  
DB 141 TCCATAATTCATATACACAGACCAATTCCTGCGCATTCACATTTTAAATTAATGAGAA 82  
|||||

QY 121 GTGATAGAGAACTGACGTGTCATATAGCTAGGGCTGAAATTTTGTGCAGATAAATAA 180  
|||||  
DB 81 GTGATAGAGAACTGACGTGTCATATAGCTAGGGCTGAAATTTTGTGCAGATAAATAA 22  
|||||

QY 181 ATAAATCATTCATCTTTT 199  
|||||  
DB 21 ATAAATCATTCATCTTTT 3  
|||||

RESULT 6  
AI721275/c 501 bp mRNA EST 10-JUN-1999  
LOCUS as82h08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
DEFINITION IMAGE:2353263 3' similar to SW:ECIC\_BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.

ACCESSION AI721275  
NID 95038531  
VERSION AI721275.1 GI:5038531  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
AUTHORS 1 (bases 1 to 501)  
Hillier L., Allen M., Bowles L., Dubuque T., Gelsel G., Jost S.,  
Kritzman D., Kuwaba T., Lacy M., Le N., Lennon G., Marra M.,  
Martin J., Moore B., Schellenberg K., Steptoe M., Tan F.,  
Teisling B., White Y., Wylie T., Waterston R. and Wilson R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246918.

CONTACT: Wilson R.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -400P from Glibco  
High quality sequence stop: 394.  
Location/Qualifiers  
1..501  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="xp11.2"  
/clone="IMAGE:2335263"  
/clone.lib="Barstead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCTGAGTGGAGCGCGCCGCTGCTGATTTTAAATTAATGAGAA  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[5' ATTCTAGTATAT 3' and 5' ATTCTAGTG 3'], digested  
with Not I and cloned into the Not I and Eco RI sites of  
the modified p773 vector. Library constructed by Bob  
Barstead."

BASE COUNT 158 a 83 c 106 g 154 t

Query Match 80.4%; Score 194.6; DB 51; Length 501;  
Best Local Similarity 95.7%; Pred. No. 1.9e-34;  
Matches 200; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GTTATTTCTCCACAGACCTCCGCGCAGACACCTAGTCTGTGTAAGAAAGCTGCTCTCTG 60  
|||||  
DB 214 GTTATTTCTCCACAGACCTCCGCGCAGACACCTAGTCTGTGTAAGAAAGCTGCTCTCTG 155  
Matches 200; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 61 TCCATAATTCATATACACAGACCAATTCCTGCGCATTCACATTTTAAATTAATGAGAA 120  
|||||  
DB 154 TCCATAATTCATATACACAGACCAATTCCTGCGCATTCACATTTTAAATTAATGAGAA 95  
|||||

QY 121 GTGATAGAGAACTGACGTGTCATATAGCTAGGGCTGAAATTTTGTGCAGATAAATAA 180  
|||||  
DB 94 GTGATAGAGAACTGACGTGTCATATAGCTAGGGCTGAAATTTTGTGCAGATAAATAA 35  
|||||

QY 181 ATAAATCATTCATCTTTT 209  
|||||  
DB 34 ATAAATCATTCATCTTTT 6  
|||||

RESULT 7  
AI721121/c 335 bp mRNA EST 10-JUN-1999  
LOCUS as73d08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
DEFINITION IMAGE:2334351 3' similar to TR:088826 O88826 GOB-5 PROTEIN. ; mRNA

sequence.  
ACCESSION AIR21121  
NID 95038377  
VERSION AIR21121.1 GI:5038377  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 335)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Getzel, G., Jost, S.,  
Kritman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
On Jun 22, 1998 this sequence version replaced gi:3246762.  
TITLE JOURNAL  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from Gibco  
High quality sequence stop: 312.

FEATURES  
source

1.335  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2334351"  
/clone\_lib="Barstead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site:1: EcoRI; Site:2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACGATCTGAGTGGAGCGGCCGCCCTTTTCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[5' ATTCACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested  
with Not I and cloned into the Not I and Eco RI sites of  
the modified pT73 vector. Library constructed by Bob  
Barstead."

BASE COUNT 98 a 57 c 71 g 109 t  
ORIGIN

Query Match 76.0%; Score 183.8; DB 51; Length 335;  
Best Local Similarity 96.4%; Pred. No. 5.2e-32;  
Matches 188; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GTTATTCCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAACGCTGCTCCTTG 60  
DB 195 GTTATTCCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAACGCTGCTCCTTG 136  
OY 61 TCCTAATATTCATATCAACAGACCATTCCTGCGATTCACATTTTAAATATATGCGAA 120  
DB 135 TCCTAATATTCATATCAACAGACCATTCCTGCGATTCACATTTTAAATATATGCGAA 76  
OY 121 GTGATAGAGAACTGACCTGTCAATAGCCTTAGGCGCTGAATTTTTCAGATTAATAA 180  
DB 75 GTGATAGAGAACTGACCTGTCAATAGCCTTAGGCGCGAAATTTCCATCAGATTAATAA 16  
OY 181 ATTAATCATTCATCC 195  
DB 15 ATTAATCATTCATCC 1

RESULT 8

AA296955  
LOCUS AA296955 378 bp mRNA EST 18-APR-1997  
DEFINITION EST112726 Colon I Homo sapiens cDNA 5' end, mRNA sequence.  
ACCESSION AA296955  
NID 91949515  
VERSION AA296955.1 GI:1949515  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
AUTHORS

1 (bases 1 to 378)  
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulmer, R.A.,  
Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,  
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C.,  
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,  
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S.,  
Glodex, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,  
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudex, D.M., Shirley, R.,  
Small, R.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,  
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
Dinke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,  
He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,  
Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H.,  
Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M.,  
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,  
Fraser, C.M., and Venter, J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl.), 3-174 (1995)  
96026280  
On Sep 12, 1996 this sequence version replaced gi:1288227.  
Other ESTs: THC167738

JOURNAL  
MEDLINE  
COMMENT

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423

Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tldb/hg1/hg1.html>)  
Seq primer: M13 Reverse.

FEATURES  
source

1.378  
/organism="Homo sapiens"  
/db\_xref="ATCC (Inhost):127834"  
/db\_xref="taxon:9606"  
/clone\_lib="Colon I"  
/dev\_stage="adult"  
/note="Organ: colon; Vector: pBluescript SK-; Site:1:  
EcoRI; Site:2: XhoI"

BASE COUNT 112 a 79 c 64 g 119 t 4 others  
ORIGIN

Query Match 74.8%; Score 181; DB 31; Length 378;  
Best Local Similarity 94.7%; Pred. No. 2.1e-31;  
Matches 196; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

OY 1 GTTATTCCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAACGCTGCTCCTTG 60  
DB 174 GTTATTCCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAACGCTGCTCCTTG 233  
OY 61 TCCTAATATTCATATCAACAGACCATTCCTGCGATTCACATTTTAAATATATGCGAA 120  
DB 234 TCCTAATATTCATATCAACAGACCATTCCTGCGATTCACATTTTAAATATATGCGAA 293  
OY 121 GTGATAGAGAACTGACCTGTCAATAGCCTTAGGCGCTGAATTTTTCAGATTAATAA 180  
DB 15 ATTAATCATTCATCC 1





mouse@eastwson.royal.tgdu  
 is available royalty-free through LLNL; contact the  
 consortium (info@image.llnl.gov) for further information.  
 34  
 iter: -28ml3 rev2 ET from Amersham  
 ility sequence stop: 374.  
 Location/Qualifiers  
 1..386  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_1="IMAGE:1167422"  
 /clone\_lib="barcoded mouse irradiated colon MFLR87"  
 /dev\_stage="8 weeks"  
 /lab\_host="Dh10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained

```

FEATURES      High quality sequence stop: 440.
               Location/Qualifiers
               1..482
               /organism="Mus musculus"
               /strain="FVB/N"

```

```

BASE COUNT      133 a      123 c      105 g      121 t
ORIGIN
/db_xref="taxon:10090"
/map="17q21"
/clone="IMAGE:1176500"
/clone_lib="Barstead mouse irradiated colon MPlRB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
15'GTATCGAATCTGAAGTGACGCGCCGCCCTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors (AATTCGATCCTGG), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library constructed by Bob Barstead."

```

[illegible]

Oy 81 GCACATTCCTCGGCAATTCACATTTTAAAAATATGTGGAATGGAAATGGAGC 140  
 Db 299 GCACATTCCTCGGCAATTCACATTCGTGAAGAATAATGTGAATGGCTAGGGGAATGGAGG 358  
 Oy 141 TGTCAATAGCCTAGGCGCTGAATTTTGTGACAGTAATAATTAATCATCATCCTTTT 200  
 Db 359 TGACCTAGTGCTTGCACATGAATTTTCAGGCAAGAAATACCA-----GTATTCCT 410  
 Oy 201 TTTCATTTAAAAATTTTCTAAAAATGATTTAGACTTCCTCT 242  
 Db 411 TTCACCTGGAGAAATTTCTAAAAAATGTACTTTAGACTTCCTGT 452

RESULT	15
AA733415	
LOCUS	
DEFINITION	AA733415 488 bp mRNA EST 07-JAN-1998
ACCESSION	v75db06.r1 Barstweed mouse irradiated colon MLR2B7 Mus musculus CDNNAL
NID	c10ne IMAGE:1176547 5' similar to SW:ECCL_BOVIN P54281 EPITHELIAL
VERSION	CHLORIDE CHANNEL PROTEIN ; mRNA sequence.
KEYWORDS	AA733415
SOURCE	g2755082
ORGANISM	AA733415.1 GI:2755082
REFERENCE	EST.
AUTHORS	house mouse.
TITLE	Mus musculus
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
COMMENT	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 488) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelsing,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R. The WashU-HMNI Mouse EST Project Unpublished (1996)

Contact: Maria M/Mouse EST Project  
MashU-HNMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@wustl.edu](mailto:mouseest@wustl.edu)  
This clone is available royalty-free through LINL ; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.

**FEATURES**  
 MGI:634795  
 Putative full length read  
 vector to vector length is 542  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 486.  
 location/Qualifiers  
 1..488 - Two nucleotides

BASE COUNT 133 a 123 c 108 g 124 t  
ORIGIN

Query Match	38.0%	Score 92:	DB 37:	Length 488;
Best Local Similarity	67.1%	Pred. No. 1.4e-11;		
Matches 149;	Conservative 0;	Mismatches 65;	Indels 8;	Gaps 1:
QY	21	GGCCAGACACACTAGTCCGTATGAAGCTGCTCCTGTGCTCTAATATTCATATCAACA	80	
Db	235	CTCAGAGACCCGCCCTATCCCGAAGACTCAATCCCCCTGTGCTGACTACAGATCAACA	294	
QY	81	GCACATTCCTCGGCATTCACATTTTAAAAATATGTGGAAGGATATAGAGAACTGCACG	140	
Db	235	GCACATTCCTCGGCATTCACAGTCTGAAGATATATGTGGAAGGCTATAGGGGAAATGCACG	354	
QY	141	TGTCATATGACCTAGGCGCTGAATTTTGTGCAGATAATATTAATTCATTCATCCCTTTT	200	
Db	355	TGACACTAGGTGTTCGACTGAATTTTCAGGCAAGAATCAACA-----GTCAATTCCT	406	
QY	201	TTTGATTTAAAAATTTCTAAATATCTATTTAGATCTTCTGT	242	
Db	407	TTCACTGGAAATTTCTTAAAAATATCTATTTAGATCTTCTGT	448	

Search completed: August 6, 1999, 13:33:29  
Job time: 13007 sec

---

This Page Blank (uspto)



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:14:38 ; Search time 3264.21 Seconds  
(without alignments)  
150.042 Million cell updates/sec

Title: US-09-049-696-15  
Perfect score: 154  
Sequence: 1 ACAGATCTTTCATGTGTAAT.....TCCTTGTCCTAATATTCATA 154

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:★

1:	gb_bal:	*
2:	gb_bal2:	*
3:	gb_om:	*
4:	gb_ov:	*
5:	gb_pat:	*
6:	gb_ph:	*
7:	gb_p11:	*
8:	gb_p12:	*
9:	gb_prl:	*
10:	gb_pr2:	*
11:	gb_pr3:	*
12:	gb_ro:	*
13:	gb_st:	*
14:	gb_sts:	*
15:	gb_sy:	*
16:	gb_un:	*
17:	gb_v1:	*
18:	em_fun:	*
19:	em_hlg:	*
20:	em_hum1:	*
21:	em_hum2:	*
22:	em_in:	*
23:	em_om:	*
24:	em_or:	*
25:	em_ov:	*
26:	em_pat:	*
27:	em_ph:	*
28:	em_p1:	*
29:	em_ro:	*
30:	em_sts:	*
31:	em_sy:	*
32:	em_un:	*
33:	em_v1:	*
34:	gb_hlg1:	*
35:	gb_hlg2:	*
36:	gb_in1:	*
37:	gb_in2:	*
38:	em_bal:	*
39:	em_bal2:	*
40:	em_hum1:	*
41:	em_hum2:	*
42:	gb_pr4:	*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	154	100.0	3311	11	AF039400	Homo sapi

2	154	100.0	2826	42	AF127036
3	152.4	99.0	35278	11	AF039401
4	143	92.9	878	5	I95746
5	73.6	29.8	2937	12	AB011756
6	47.9	29.9	3415	42	AF043976
7	46	29.8	3415	12	AF043976
8	35.6	23.1	2984	3	BTU36445
9	35	22.7	3471	12	AF052746
10	34.6	22.5	23091	37	CELT733H2
11	33.4	22.3	3604	9	AB026633
12	32.2	20.9	3022	12	AF04838
13	32.2	20.9	15158	34	HSDJ3994
14	32	20.8	99083	36	CEY3719B
15	31.6	20.5	8397	1	BSU55043
16	31.6	20.5	3317	3	AF001261
17	31.6	20.5	123189	11	AC002556
18	31.4	20.4	137137	42	AC004928
19	31.4	20.4	100258	11	AC000746
20	31.4	20.4	267	14	HSB060X5
21	31	20.1	192645	34	AC006673
22	31	20.1	1111	3	PtG1YSOZYM
23	31	20.1	12975	3	SSU28757
24	30.8	20.0	3961	7	SACREIDNA
25	30.8	20.0	5420	7	YSCGDCPK
26	30.8	20.0	39203	8	YSCH9205
27	30.8	20.0	29468	9	HS431H6
28	30.8	20.0	54336	11	HUMXPDG1
29	30.8	20.0	2457	36	AB009875
30	30.4	19.7	36555	37	CELCA38C3
31	30.2	19.6	20891	36	CEFO0136
32	30	19.5	12206	2	AF000366
33	30	19.5	1301	36	SEVAP1
34	29.8	19.4	56458	9	HS432B11
35	29.8	19.4	814	11	HSXNP14
36	29.8	19.4	80865	35	AC004378
37	29.8	19.4	26325	35	CER00541
38	29.6	19.4	124700	37	AC005558
39	29.6	19.2	12772	37	CELF53A10
40	29.4	19.1	3213	12	AF104414
41	29.4	19.1	173839	35	AC007483
42	29.4	19.1	74534	37	AC004735
43	29.4	19.1	4307	42	AF104413
44	29.2	19.0	11360	1	AB001721
45	29.2	19.0	19243	1	AB010203
					AB010203 Leptospir
					AF127036 Homo sapi
					AF039401 Homo sapi
					155746 Sequence 8
					AB011756 Mus musc
					AF043976 Homo sapi
					156445 Bos taurus
					AF052746 Mus musc
					U80033 Caenorhabdt
					AB026633 Homo sapi
					AF04838 Mus musc
					AL078595 Homo sapi
					AL02835 Caenorhab
					U55043 Bacillus su
					AF001261 Bos tauru
					AC002556 Human Chr
					AC004928 Homo sapi
					AC000746 Human BAC
					253246 H.sapiens (
					AC006763 Caenorhab
					L14853 Sus scrofa
					U28757 Sus scrofa
					U44435 Sus scrofa
					L19670 Saccharomyc
					U15540 Saccharomyc
					U1055 Saccharomyc
					AL03109 Human DNA
					L47234 Homo sapien
					AB009875 Octopus d
					AF024494 Caenorhab
					268341 Caenorhabdt
					AD000366 Borrelia
					X69993 S.purpuratru
					AL008719 Human DNA
					U97085 Homo sapien
					AC004378 Drosophil
					275546 Caenorhabdt
					AC005558 Drosophil
					AF016660 Caenorhab
					AF

RESULT	1
AF039400	
LOCUS	
DEFINITION	AF039400 3311 bp mRNA
ACCESSION	Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,
NID	AF039400 complete cds.
VERSION	AF039400.1 GI:4009457
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 3311)
JOURNAL	Gruber,A.D., Ehble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and
MEDLINE	Pauli,B.U.
REPERE	Genomic cloning, molecular characterization, and functional
AUTHORS	analysis of human CLCA1, the first human member of the family of
TITLE	Ca <sup>2+</sup> -activated Cl- channel proteins
	Genomics 54 (2), 200-214 (1998)
	99047526
	2 (bases 1 to 3311)
	Gruber,A.D., Ehble,R. and Pauli,B.U.
	Direct Submission

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers

source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue-type="small intestine"  
1. .3311  
/gene="hCAL1"  
352. .3096  
/gene="hCAL1"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPFKSSVFLLILHLLEGALNSLIQLNNNGEIVADPNP  
EDETLIQIKDMVTQASLYLEATGRFEYFNVAIILPETWTKADYVRPLETKNA  
DVLVAESPENDEPYTEOMNGCEGGERHILPDLTAKGLAEYPOGAFFHEMAH  
LRMGVDEYNDDEKFTLSNGRIQAVRSAGITGTNVYKKCGGSCYTRKCTFENVTGL  
YEGCEVYLSRQTEKASIMFAOHVDSYVECTEONHNKRNPNKONKCNLRBTWEVI  
RQSEDFKTTPTPTOPPPNPTFSLQIGRIVCLVLDKSGMATGNRLNRLNOGQLFL  
LQVLEKSWGMVTFDSAAHVSLELIQINSGRDYLAKRLPAASGGTSLCSGLNSA  
FTVIRKRYPTDGESEVILTDGENDTISGCENEVQSGAIHTVALGPSAOLELSEK  
MTGLOIYASDOVONNGILDAFGALSSGNVAVSORSIOLESKGLTLONSOMNGTVIV  
DSVAGDITLITWTQPOPLLMDPSGOGGFEVDKNTKMAVLIQIPGAKVGTWKY  
SLAASSQITLITLTSRASNAITLPITVTSKNTKDSKPSPLVYVANITOGASPIIRA  
SVTALLESVNGKTVTLELDNGAGADNRKDGYSRFTTYDNGRSYKVRALGYN  
AARRVYIPQSGALYIPGWIENDEIQMNPPEINKDDVOHKVCSRTSSGGSFVAS  
DVPNAPIPDLFPPGOITDLKAEIHGSLINLTWAPGDYDHTAKHYIIRISTSLID  
LRKFNEISLQVNTTALIPKEANSEVEFLFENITFENGDTLFTAIQAVDKVLKSEI  
SNTARVSLFIPQTPPETPSPDETSAFCPNHINSTIPGIIHLKIMMKWIGELQSLIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 154; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 7.6e-36;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACAGATCTTTTCATTCGATTCAGGCTGTTGATTAAGTGCATCGAATCAGAAATATCC 60  
|||||  
DB 2863 ACAGATCTTTTCATTCGATTCAGGCTGTTGATTAAGTGCATCGAATCAGAAATATCC 2922  
OY 61 AACATTGCACGAGTATCTTTGTTTATCTCTCCACAGACTCCGCCAGACACCTAGTCT 120  
|||||  
DB 2923 AACATTGCACGAGTATCTTTGTTTATCTCTCCACAGACTCCGCCAGACACCTAGTCT 2982  
OY 121 GATGAAGCTGCTGCTTGTCTTAATATTCATA 154  
|||||  
DB 2983 GATGAAGCTGCTGCTTGTCTTAATATTCATA 3016

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2826)  
AUTHORS Agnel, M., Vernat, T. and Culouscou, J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride

JOURNAL channel  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel, M. and Culouscou, J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Neuilly-Malmaison 92500, France  
FEATURES Location/Qualifiers

source  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue-type="small intestine; colon"  
1. .2826  
/gene="CACCL1"  
5. .2749  
/gene="CACCL1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AA025487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPFKSSVFLLILHLLEGALNSLIQLNNNGEIVADPNP  
EDETLIQIKDMVTQASLYLEATGRFEYFNVAIILPETWTKADYVRPLETKNA  
DVLVAESPENDEPYTEOMNGCEGGERHILPDLTAKGLAEYPOGAFFHEMAH  
LRMGVDEYNDDEKFTLSNGRIQAVRSAGITGTNVYKKCGGSCYTRKCTFENVTGL  
YEGCEVYLSRQTEKASIMFAOHVDSYVECTEONHNKRNPNKONKCNLRBTWEVI  
RQSEDFKTTPTPTOPPPNPTFSLQIGRIVCLVLDKSGMATGNRLNRLNOGQLFL  
LQVLEKSWGMVTFDSAAHVSLELIQINSGRDYLAKRLPAASGGTSLCSGLNSA  
FTVIRKRYPTDGESEVILTDGENDTISGCENEVQSGAIHTVALGPSAOLELSEK  
MTGLOIYASDOVONNGILDAFGALSSGNVAVSORSIOLESKGLTLONSOMNGTVIV  
DSVAGDITLITWTQPOPLLMDPSGOGGFEVDKNTKMAVLIQIPGAKVGTWKY  
SLAASSQITLITLTSRASNAITLPITVTSKNTKDSKPSPLVYVANITOGASPIIRA  
SVTALLESVNGKTVTLELDNGAGADNRKDGYSRFTTYDNGRSYKVRALGYN  
AARRVYIPQSGALYIPGWIENDEIQMNPPEINKDDVOHKVCSRTSSGGSFVAS  
DVPNAPIPDLFPPGOITDLKAEIHGSLINLTWAPGDYDHTAKHYIIRISTSLID  
LRKFNEISLQVNTTALIPKEANSEVEFLFENITFENGDTLFTAIQAVDKVLKSEI  
SNTARVSLFIPQTPPETPSPDETSAFCPNHINSTIPGIIHLKIMMKWIGELQSLIA  
"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 154; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 7.5e-36;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACAGATCTTTTCATTCGATTCAGGCTGTTGATTAAGTGCATCGAATCAGAAATATCC 60  
|||||  
DB 2516 ACAGATCTTTTCATTCGATTCAGGCTGTTGATTAAGTGCATCGAATCAGAAATATCC 2575  
OY 61 AACATTGCACGAGTATCTTTGTTTATCTCTCCACAGACTCCGCCAGACACCTAGTCT 120  
|||||  
DB 2576 AACATTGCACGAGTATCTTTGTTTATCTCTCCACAGACTCCGCCAGACACCTAGTCT 2635  
OY 121 GATGAAGCTGCTGCTTGTCTTAATATTCATA 154  
|||||  
DB 2636 GATGAAGCTGCTGCTTGTCTTAATATTCATA 2669

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCAL1) gene,  
DEFINITION complete cds.  
ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source 1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="11 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PDB:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPKSSVFILILHLLEGALNSLIQLNNNGEIVVAIDPNV EDELIIQIKDWATQASLYLFATGKREFKRVALLPEPTKADYEPKLEYKNA DIVLAVESTPGNDEPTEGMNGEKGRIHILPPFIAGKRIAEYPOGKAVHMAH LRMGVDEYNNDEKFTYNGRFOARCSAGITGTNVVKKCGGSCYTRKCTPNKTYGL YKGEFEVQSRQTEKASTMFQAHVDSTVERTEQHNKEAPNKKONKLNSTWEV L RDSDFKRTPTPTOPNPPTFSLQIGRIVCLVLDKSGMATGNRLNLNQAGOLF LQTVELGSMVWAVTDPDAHVQSELQIINSGSDRDTLAKRLPAASGGSTISGSRSA FTVIRKKYPTDSEIVLILDGEDNTISGCFNKGSGALIHVALAPSAQGLELSK MTGGLQTVASDOVONGLIDAGALSIGGFANVSOSIOLSGSLTQNSOMNGVIV DSTVGKRLFLITVTTOPPOQILLMPDSGCKGCFVADNRKAYIQIPIRIAVGWKY SIQASSQTLITVTSTRASNATLPITTVTSKIDKSKFSPPLVATANIQSGAPLRA SVTALIESVNGKTVLELDNGAGADATDDVYSRYFTYDNGRYSVKVALGQVNA AARRVPIQOOSGALYIPGWIENDEIQWNPPEIKNDQVQHVCFSTRSGSFEVAS DVPNAPIDLPFGQITDLNAEIHGSLNFWTAPGDYDHGTAKHYIRISTSID LRDKNESLOVNTALTILPEANSEVFLFKPENITFENGTDLFIAIQADVDKLSEI SNIAVNSLFIPTQTPETPSPDSTAPCNININISTIPQIHILKIMKMGISLQISA "

exon 6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"

exon 6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"

exon 9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15

exon 9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15

exon 9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15

BASE COUNT 11245 a 6870 c 7230 g 9933 t

ORIGIN

Query Match 99.0%; Score 152.4; DB 11; Length 35278;  
Best Local Similarity 99.4%; Pred. No. 2.5e-35;  
Matches 153; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTATCTTCTATTCAGCTGTGTGATTAAGTCGATCTGAATCAGAAATATCC 60  
|||||  
DB 33077 ACAGATCTTTTATCTTCTATTCAGCTGTGTGATTAAGTCGATCTGAATCAGAAATATCC 33136  
|||||

QY 61 AACATTGACAGAGTATCTTGTATTCTCCACAGACTCCGACAGACACTAGTCT 120  
|||||  
DB 33137 AACATTGACAGAGTATCTTGTATTCTCCACAGACTCCGACAGACACTAGTCTCC 33196  
|||||

QY 121 GATGAACGCTGCTCTCTGCTCTAATATTCATA 154  
|||||  
DB 33197 GATGAACGCTGCTCTCTGCTCTAATATTCATA 33230  
|||||

RESULT 4  
LOCUS 195746 878 bp DNA PAT 17-JUL-1998  
DEFINITION Sequence 8 from patent US 5733748.  
ACCESSION 195746  
NID 93940216  
VERSION 195746.1 GI:3940216  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 878)  
AUTHORS Yu,S. and Rosen,C.  
TITLE Colon specific genes and proteins  
JOURNAL Patent: US 5733748-A 8 31-MAR-1998;  
FEATURES Location/Qualifiers  
source 1..878  
/organism="unknown"

BASE COUNT 257 a 179 c 188 g 241 t 13 others  
ORIGIN  
Query Match 92.9%; Score 143; DB 5; Length 878;  
Best Local Similarity 99.4%; Pred. No. 1.3e-32;  
Matches 154; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
OY 1 ACAGATCTTTTCATTCATTCAGCTGTGATAGTCGATCGAATATATATCC 60  
Db 545 ACAGATCTTTTCATTCATTCAGCTGTGATAGTCGATCGAATATATATCC 604  
OY 61 AACATTCGACGAGTATCTTTGTTATTCCTCCACAGACTCCGACAGACACTAGTCTT 120  
Db 605 AACATTCGACGAGTATCTTTGTTATTCCTCCACAGACTCCGACAGACACTAGTCTT 664  
OY 121 GATGAACGCTGCTGCTGCTGT-CTTAATATTCATA 154  
Db 665 GATGAACGCTGCTGCTGCTGCTTAATATTCATA 699  
RESULT 5  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS AB017156 Mus musculus gob-5 mRNA, complete cds.  
DEFINITION AB017156  
ACCESSION 93721911  
VERSION AB017156.1 GI:3721911  
KEYWORDS GOB-5.  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (sites), Tanigawa, Y. and Hirohashi, S.  
Cloning and identification of the gene gob-5, which is expressed in  
intestinal goblet cells in mice  
Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
JOURNAL 2 (bases 1 to 2937)  
MEDLINE 99160866  
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru  
Komiyama, ERA70, Jst, Hirohashi Cell Configuration Project; 5-9-9,  
Tokodai, Tsukuba, Ibaraki 300-2635, Japan  
(E-mail:tkom@ccp.jst.go.jp, Tel:81-298-47-7563,  
Fax:81-298-47-5226)  
FEATURES  
source Location/Qualifiers  
1..2937  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15..2756  
/gene="gob-5"  
15..2756  
/gene="gob-5"  
/codon\_start=1  
/protein\_id="BA033743.1"  
/db\_xref="PID:di034712"  
/db\_xref="PID:93721912"  
/db\_xref="GI:3721912"  
/translation="MESLSPVFLILHLIEGLVSLTLOLNNGNIEGIVIAIDHVP  
/EDALIOHIDKMTQASPLYFEATGRFFKFNVAIIPSWAKPPTPKLETKNA  
DVLVSTSPGNDPEYTERHAGCEGKIRHLTPDLAKKLTQYGPDPRTYVWAH  
FRGCVENVNDKEFKYLSKGPQAVKCSAITKNOVRCOGSCITNGKCVIDRVG  
LYKNDCTVFDPRHONKASTNENINSVVECTEKNNHONADNDONRNLSTIEV  
IOSEDDKQTPMTAOPAPATFSLIOGATVIGVLDKSGMLNDRLNRMNDASLE  
LQTEDEGSMVAVTDESAIVVSELQKLSGADRDLILKHLFTYVAGAGTSCIGRT  
AFYVKKRYPTDSEIVLITDGEDNTISSCFDLVKSGALITHVALGAPAAKLEBLS  
KMTGGLOTYSDDVONNGILVDAFALSSGNAAIAOHSIOLESNGVNLONNOMNGSVI  
VDSVVCMDTFLITWTHTHPTLIFIMDSVGEONGFILDITTKAVYLQVGTAKVGMK

BASE COUNT 860 a 718 c 693 g 666 t  
ORIGIN  
Query Match 47.8%; Score 73.6; DB 12; Length 2937;  
Best Local Similarity 81.7%; Pred. No. 4.1e-12;  
Matches 85; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
OY 1 ACAGATCTTTTCATTCATTCAGCTGTGATAGTCGATCGAATATATATCC 60  
Db 2526 ACAGATCTTTTCATTCATTCAGCTGTGATAGTCGATCGAATATATATCC 2585  
OY 61 AACATTCGACGAGTATCTTTGTTATTCCTCCACAGACTCCGCC 104  
Db 2586 AACATTCGACGAGTATCTTTGTTATTCCTCCACAGACTCCGCC 2629  
RESULT 6  
AF043976 3415 bp mRNA PRI 07-APR-1999  
LOCUS AF043976 Homo sapiens CLCA homolog (hCLCA3) mRNA, complete cds.  
DEFINITION AF043976  
ACCESSION 94572288  
VERSION AF043976.1 GI:4572288  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 3415)  
AUTHORS Gruber, A.D. and Pauli, B.U.  
TITLE Molecular cloning and biochemical characterization of a truncated,  
secreted member of the human family of Ca2+-activated Cl- channels  
Biochim. Biophys. Acta 1444 (3), 418-423 (1999)  
JOURNAL 2 (bases 1 to 3415)  
MEDLINE 99196715  
AUTHORS Gruber, A.D., Eble, R.C. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source Location/Qualifiers  
1..3415  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..3415  
/gene="hCLCA3"  
19..807  
/gene="hCLCA3"  
/codon\_start=1  
/product="CLCA homolog"  
/protein\_id="AAD33734.1"  
/db\_xref="PID:94572289"  
/db\_xref="GI:4572289"  
/translation="MVFSLKYLFLSLSPVLKSLVLTNNNGYDGVIAINPSYE  
DEKLTQNKKEVTEASTHTLPATKORAVFNVSLIIMPTKSKSEYLIPOEYDQAD  
VIVADLVKIGDDPTLQIGCGCKGQYIHTPTFLIINNLTATGPRGKVFVGMAL  
RMGVFDEYNDQPYISRNRTMTQSTRITIVMTNECKGASCIARPPRRSQICL  
YEACKEFIPKRSQYAKSISIVFMONIDSVTERCTEKTINKENPNL"  
BASE COUNT 1149 a 625 c 630 g 1011 t  
ORIGIN  
Query Match 29.9%; Score 46; DB 42; Length 3415;  
Best Local Similarity 68.1%; Pred. No. 0.00058;  
Matches 64; Conservative 0; Mismatches 30; Indels 0; Gaps 0;  
OY 8 TTTTCATTCGATTCAGCTGTGATAGTCGATCGAATATATATCCACATTG 67

Db 2609 TCTATATTCATTCATTCAGCATTCAGAGCCATGTCACCTGAGGTTTCAACATTTG 2668  
Qy 68 CACGAGTATCTTTGTTTATTCCTCCACAGACATCC 101  
Db 2669 CACAGCACTATCTATTCCTCCACAGAAC 2702

RESULT 7  
BTU36445 2984 bp mRNA MAM 09-FEB-1996  
LOCUS Bos taurus calcium-activated chloride channel mRNA, complete cds.  
DEFINITION U36445  
ACCESSION 91184065  
VERSION 036445.1 GI:1184065  
KEYWORDS  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Bovidae; Bos.  
REFERENCE 1 (bases 1 to 2984)  
Cunningham, S.A., Awayda, M.S., Bubien, J.K., Ismailov, I.I.,  
Arrate, M.P., Berdiev, B.K., Benos, D.J. and Fuller, C.M.  
Cloning of an epithelial chloride channel from bovine trachea  
J. Biol. Chem. 270 (52), 31016-31026 (1995)  
JOURNAL 96125078  
MEDLINE 2 (bases 1 to 2984)  
REFERENCE Benos, D.J.  
AUTHORS Direct Submission  
TITLE Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,  
University of Alabama at Birmingham, BHSB 706, Birmingham, AL  
JOURNAL 35294, USA

FEATURES  
source  
1. 2984  
Location/Qualifiers  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19..2730  
/note="epithelial chloride channel"  
/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="pid:91184066"  
/db\_xref="gi:1184066"  
/translation="MVPRLTVILFLTLHLPLPKMSMYNLINNGYGVIVIAINPSVPE  
DEKLONIKENVTASTYLPHATKRRVFRVNSILIPMTKSKSEYLMKPOSSVOAE  
VIVANPYLKHGDDPTTLOYGRGEGQYIHTPNFLINNLPIYSGRGAIPHVAHL  
RMGLIDENGDPTTISRNNTIATRCSTHITGIVAVCOGSCITPRCRDSOTGL  
KMGATIDENGDPTTISRNNTIATRCSTHITGIVAVCOGSCITPRCRDSOTGL  
YEACTIPEKQATRESIMFQSLHVECTEKTHTVAVENALONKNCNGSTWDI  
MNSDFONTPTEEMNPPTOPFSLKSKORVCLYLDKSGMSSEDLFRNQAAEL  
FLIITIEKSLVGMVTPDVAEIRNNILKRTIDNYENITANLPOEANGSTICGLK  
AGFOALIOSOOSTSGSEITILTDGEDNEHSGIEEVKOSGVIIHTIAGPSAKRELT  
LSDMGCHRFANDINCLTANFASISSRSGSIQTQIDESKALAIETKRVNATVP  
VDSITGNDTFFVATITKPELIDDPGKRTISDFEKDKNLSAKRLTPIGLIETG  
TWYSLNLNHAHQILIVTVTRASPPTPVTAHMSQNTAHSPTSPVIVAYAOVSQ  
FLPVGLINWTAIIERDEGHVLELMDNGAGADYKNGISRYTDPYRGNRYSKY  
HAERNNATRLSLROPNKALYIPGIEKNGIILPPREVDLAKKIEDEFSYLS  
GGSEFTVSGAPGNHPSVLPNKTIIDLEKFKRDHOLSMTAPANLDEKSNASYIRI  
SKSFIDLOKDDNATLVNTSSIKRKEASDEFEKPERFERENTNYIAVOAINEA  
NLTSVSNIAQIKFIPMPEDSVPLGTRKISAINIATRALMALTISIV"

CDS  
1027 a 583 c 813 t

BASE COUNT 1027 a 583 c 813 t  
ORIGIN

Query Match 23.1%; Score 35.6; DB 3; Length 2984;  
Best Local Similarity 62.2%; Pred. No. 0.68;  
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
Qy 1 ACAGATCTTTATCTGCTATTCAGCTGTGATTAAGTCGATCGATGAATCAGAAATATCC 60  
Db 2548 ACCAACTTATATATTCAGCTCCAGCAATCAGCAATCTACCTCAGAGGTTTCT 2607

Qy 61 AACATTCACAGATATCTTTATTCCT 90  
Db 2608 AACATTCACAGATATCTTTATTCCT 2637

RESULT 8  
AF052746 3471 bp mRNA ROD 08-APR-1999  
LOCUS Mus musculus chloride channel Cacc mRNA, complete cds.  
DEFINITION AF052746  
ACCESSION 93560546  
VERSION AF052746.1 GI:3560546  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3471)  
Romio, L., Musante, L., Cinti, R., Serl, M., Moran, O., Zegarra-Moran, O.  
and Galletta, L.J.V.  
Characterization of a murine gene homologous to the bovine Cacc  
chloride channel  
Gene 228 (1-2), 181-188 (1999)  
JOURNAL 99173882  
MEDLINE 2 (bases 1 to 3471)  
REFERENCE Romio, L., Musante, L., Cinti, R., Moran, O., Serl, M. and  
AUTHORS Galletta, L.J.V.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto  
Gianina Gaslini, Genova I-16148, Italy  
FEATURES  
source  
1. 3471  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/db\_xref="dbEST:AA245677"  
47..2752  
/codon\_start=1  
/product="chloride channel Cacc"  
/protein\_id="AAC35003.1"  
/db\_xref="pid:93560547"  
/db\_xref="gi:3560547"  
/translation="MVPGLQVILFLTLHLNLTSSMYNLINNGYGVIVIAINPSVPE  
DERLIPSIKENVTASTYLPHATKRRVFRVNSILIPMTKSKSEYLMKPOSSVOAE  
VIVANPYLKHGDDPTTLOYGRGEGQYIHTPNFLINNLPIYSGRGAIPHVAHL  
RMGLIDENGDPTTISRNNTIATRCSTHITGIVAVCOGSCITPRCRDSOTGL  
KMGATIDENGDPTTISRNNTIATRCSTHITGIVAVCOGSCITPRCRDSOTGL  
YEACTIPEKQATRESIMFQSLHVECTEKTHTVAVENALONKNCNGSTWDI  
MNSDFONTPTEEMNPPTOPFSLKSKORVCLYLDKSGMSSEDLFRNQAAEL  
FLIITIEKSLVGMVTPDVAEIRNNILKRTIDNYENITANLPOEANGSTICGLK  
AGFOALIOSOOSTSGSEITILTDGEDNEHSGIEEVKOSGVIIHTIAGPSAKRELT  
LSDMGCHRFANDINCLTANFASISSRSGSIQTQIDESKALAIETKRVNATVP  
VDSITGNDTFFVATITKPELIDDPGKRTISDFEKDKNLSAKRLTPIGLIETG  
TWYSLNLNHAHQILIVTVTRASPPTPVTAHMSQNTAHSPTSPVIVAYAOVSQ  
FLPVGLINWTAIIERDEGHVLELMDNGAGADYKNGISRYTDPYRGNRYSKY  
HAERNNATRLSLROPNKALYIPGIEKNGIILPPREVDLAKKIEDEFSYLS  
GGSEFTVSGAPGNHPSVLPNKTIIDLEKFKRDHOLSMTAPANLDEKSNASYIRI  
SKSFIDLOKDDNATLVNTSSIKRKEASDEFEKPERFERENTNYIAVOAINEA  
NLTSVSNIAQIKFIPMPEDSVPLGTRKISAINIATRALMALTISIV"

CDS  
1096 a 777 c 862 t

BASE COUNT 1096 a 777 c 862 t  
ORIGIN

Query Match 22.7%; Score 35; DB 12; Length 3471;  
Best Local Similarity 61.5%; Pred. No. 1;  
Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
Qy 5 ATCTTTATCTGCTATTCAGCTGTGATTAAGTCGATCGATGAATCAGAAATATCC 64  
Db 2574 ACCTTATCTGCTATTCAGCTGTGATTAAGTCGATCGATGAATCAGAAATATCC 2633  
Qy 65 TTGACAGATATCTTTATTCCTCCACA 95  
Db 2634 TCGCAGAGGCTGTCAAGCTTACTTCTCAGA 2664

RESULT 9  
 CELT23H2/C  
 LOCUS CELT23H2 23091 bp DNA INV 14-DEC-1998  
 DEFINITION Caenorhabditis elegans cosmid T23H2.  
 ACCESSION U80033  
 MID 92731373  
 VERSION U80033.1 GI:2731373  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Caenorhabditis elegans strain-Bristol N2.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 Rhabditina; Rhabditoidea; Rhabditidae; Pelodietinae; Caenorhabditis.  
 1 (bases 1 to 23091)  
 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,  
 Bonfield, J., Burton, J., Connell, M., Copest, T., Cooper, J.,  
 Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,  
 Fullon, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,  
 Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N.,  
 Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,  
 O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,  
 Saunders, D., Showkhen, R., Smaildon, N., Smith, A., Sonhammer, E.,  
 Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,  
 Vaughan, K., Waterston, R., Watson, A., Wellstead, L.,  
 Wilkinson-Sproat, J., and Wohlman, P.  
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans  
 Nature 368 (6466), 32-38 (1994)  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Submitted by:  
 On Dec 30, 1997 this sequence version replaced gi:1703551.

Genome Sequencing Center  
 Department of Genetics, Washington University,  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1HQ, England  
 e-mail: rwenematode.wustl.edu and jesusanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we provide a small overlap between  
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate  
 chemistry; an attempt was made to resolve all sequencing problems,  
 such as compressions and repeats; all regions were covered by  
 sequence from more than one subclone

#### NEIGHBORING COSMID INFORMATION:

The 5' cosmid is F37E3, 500 bp overlap; 3' cosmid is K06A5, 700 bp  
 overlap. Actual start of this cosmid is at base position 497 of  
 CELT23H2; actual end is at 13461 of CELK06A5

#### NOTES:

Coding sequences below are predicted from computer analysis, using  
 the program GeneFinder/P. Green and L. Hillier, ms in preparation).

FEATURES  
 source  
 1. 23091  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /chromosome="I"

gene  
 CDS  
 /clone="T23H2"  
 525..3669  
 /gene="T23H2.2"  
 join(525..594,675..793,1342..1485,1674..1973,2016..2110,  
 2405..2559,2672..2829,3538..3669)  
 /gene="T23H2.2"  
 /note="similar to the C2 domains of synaptotagmins"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="AAC48201.1"  
 /db\_xref="PID:g1703552"  
 /db\_xref="GI:1703552"  
 /translation="MFLGHSRSELLFENFENSCQLYHOMFVGVGAGVLAALVAVAA  
 LLVRRRQPSLLPQVYVAVRGSPFGPGLKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP  
 NLLEDRRLKSP  
 DPVYKQLLEPHERHVKTRIVRTSPVEEOPAMGVHEOVNFAFLHQAQVARY  
 SRDVLVGECEYRLADAELOVHNMREVELPRLPATDVARAGELLSTLYQAENILY  
 VVILKARGSLDQVADGADPEVFLYLRKEGERTVKKRTFRSTLNPVNESEVEL  
 PDRLEHSVDLDQVHNDVGRNDRVYGRALLNEDSHVEVELENPGROVAQWHLID"  
 complement(4639..9236)  
 /gene="T23H2.3"  
 complement(join(4639..4831,4886..5197,5399..6338,  
 6682..7217,7827..8314,8625..9236))  
 /gene="T23H2.3"  
 /note="strong similarity to heliases of the SNF2/RAD54  
 family; coded for by C. elegans cDNA yk43f12.3; coded for  
 by C. elegans cDNA yk134d1.3; coded for by C. elegans cDNA  
 yk53e3.5; coded for by C. elegans cDNA yk134d1.5; coded  
 for by C. elegans cDNA yk58f12.5"  
 /codon\_start=1  
 /protein\_id="AAC48202.1"  
 /db\_xref="PID:g1703553"  
 /db\_xref="GI:1703553"  
 /translation="MASFRKNHYIIDSQSSSEDEDRKVCSLTFEECFKRIEQRERS  
 SSSSEVETAMRSTIDSSSEPPRTIIDASLNMASTVITSPNSKOSISRTGSSVYGS  
 KCSIDHYRSTRSFDNSFEKPVITIKLOESDEPRAKQQLRRLRLKRSIKQSDDL  
 TGGSSRSRGCTGENSESLSLSDLSHSPIAKTFRHASDLSDMSPPNRSTLSV  
 ETSNGSNPOTPYGSP  
 DNGARKYLQVDEITEERIRLINGIDVTVDVDTTAPPPROKNIAPIDPPFAML  
 DKNGRWKSRMTQKILKYNKISDRIMQQLADVTITPTELTDPDGLVLELPH  
 OKRGLVWREGGPHSGILADMDGKTLMSLSTLVHOKAARRAKESGDNAADAE  
 KRRYAKEEGLYSGNTLILAPASLIHWEINERESDLSVFMFGTKQKQIEK  
 ELARYDVITTYTLAANELMEKKAAGKKEEDSEENENSPRPGKNDSPARVA  
 WSPYIDDEAIAIKRLSQCRAVCRTSFSFRKISGPIHNNILNLDVSLVRFPIPL  
 GDRKTAESIMPACTGMAQRYNLTIRKMLRRKDDOCALTNKIKVLRKKKIEIPL  
 EGVDEENGAIAMEAQAQVKQIVITDIDIKIGQRRRQNGDNDEFPNPVNGR  
 NLGNSNFQNMNSCILLMLRLROACVHFHTKSGMDADARKINGDEEDADMELENL  
 MEKTMADLTIDSDNDEGYDTNKEEPKQPTRFEPDFISCKNAKTLLEVRLTEK  
 EKVIYVQMTSVNLVLEKHIOAGGHNTSITGVILVROEROERSEFNOEGAGQVML  
 SITRAGGVGLNLIGSNHLIMVDLHMNPALDEQACDRIVRMGOKKEVFTRHNSIFCR  
 IYKCTIQRVMDLDEKTLAALASVLEGTARLKLKLTATIRMLFGDGSNMN"  
 10104..16369  
 /gene="T23H2.1"  
 join(10104..10258,10307..10455,10498..10862,10911..11172,  
 11219..11397,11464..12747,12798..13722,13772..14141,  
 14217..14373,14420..14598,14666..15183,15233..15561,  
 15605..15920,15966..16190,16239..16369)  
 /gene="T23H2.1"  
 /note="strong similarity to rat integral membrane  
 glycoprotein gp120 precursor (SP:P11654); coded for by C.  
 elegans cDNA yk159h8.5; coded for by C. elegans cDNA  
 yk179g4.5; coded for by C. elegans cDNA yk17b3.3; coded  
 for by C. elegans cDNA yk17b3.5; coded for by C. elegans  
 cDNA yk28d6.5; coded for by C. elegans cDNA yk34c9.3;  
 coded for by C. elegans cDNA yk34c9.5; coded for by C.  
 elegans cDNA yk43e1.3; coded for by C. elegans cDNA  
 yk43e1.5; coded for by C. elegans cDNA yk48f10.3; coded  
 for by C. elegans cDNA yk130e9.3; coded for by C. elegans  
 cDNA yk48f10.5; coded for by C. elegans cDNA yk53b6.5;  
 coded for by C. elegans cDNA yk130e9.5"  
 /codon\_start=1  
 /protein\_id="AAC48199.1"  
 /db\_xref="PID:g1703554"

/db\_xref="GI:1703554"  
/translation="MILRSLLVLAIVOTITISRLNVPVLLPYHPTVPSFLVETH  
PBGCGFTRSTPDIYSKRITNAGCDRAKAPISVAKPGVSGSELSAVFADKQ  
SGTLLSCGYTVDIATISTEITTKVLEVDAAARNTVDNFADGDFSTLSIALEME  
LASTSSNNAKPLRIPEEQSTYEAPSEIKYKLEKNADGKGLLIEVGTGATLTKRS  
DAYLQKVAHNEVLAVANLLVPSODVLPVHVSYPVLLVQKQGEIVAMPNPSY  
EALDQGEVLAHLKSSSVRALTIGNTAVHLSHVDYAKGLRPPSTVIVHVS  
VOMHYSGMMLETKOYINVEILDEHGNWAFVADNSPDHIDEPHVDKESNG  
TPELTPKPSKTTLRKFAVLIIDAKGNRIASQSGQRTIYDPRIVPYLLP  
FVSEKRSQIDLTATGSGGLEWTSDEGHAIVDLTIGRTANSLSSTKAKDKNDO  
LMDIASVHLEVSIGFETVRETVGDTLLNINAKGTLSGLVEMDCIDVIAVAY  
OTDNALLHRESSADSLPMGTGCTITFKLSGDAVSSTYGHKASIDVAYEK  
LSISESSSIALGTHPLTVSGGPRMFLDPANFKTOETKOSLOVPEKRVLEK  
CSSEYEAVALRIGNLKSTPLPHSETTVICCAKPRLEIPPKQKQSPCLNPL  
SHLINTVELVLRGSGVCCGATPLASINGLSPKNTTSDGCLITVRHREIDATSGK  
KEGQVTOAQSISTKYEITVKGLNVEPALVIMNEVSKGTITFGSGHFVADN  
LPTSDSPAIALRARSLLVTPKNGQVNLRIADOLVGHADAARYADINSLADAP  
OFEVIGOEVEVILLADDEGAFSEKHEHRLADQADANHHVILKRVGLARTLANS  
IGTVLSASSKSSGRLSRPHETVOIFEPILQKRLILIPDSKOLEVGVGPOTP  
PLDFSLNMSIASIENPALITSEIGYATITETVAGDHTLDTVILVAVLGGIIL  
SASSKRVETGGRVNLRIGVINGADEPFAFGALIPKRYWVSDBSVLTPTRPLG  
GDVEPTDQALFENNAIRGGSVYTKAVVELKARKHGTSTFTLETITVEDGL  
SLVQEMDINTVAPNSQLKMTASQASFSPDSRIVISADGHLITGKGSAA  
ALTIVNNSPDNEVLIPVTVSRVSLDVHPITELKSAFENPLIHLPEVQAIOQNVV  
PADARGRLAASNSINFRPHFDLTIDVAINSTQTLTTLTAGDTVLRIDASNTV  
IATFIRLSASESILPRAHKVANDLVSDVLOSNITFVDSRMSOSSSGRISML  
DENLGAOUTKAGNTFIRLHADKOTIHESITVYLSLRFPGQKPEFYNDHEVYV  
IPVIAATNTSGSKSVSYGECTAQIIRSFDAIGAPFEQVAFTRRSKLTISAVNMLTV  
SAVEPVGEGCEIRRFDSVSTSIIVEPEILKQDFARITAKMISDQVNDATV  
DYVPEHAFIVEKELEVSNNQIEALSIWAPYDSKHIVSGEGDYSVETKRS  
DKHSAKANYENIRLNKISAALETAKKCOISVENTLGOVIRVPTVOLDEAKO  
VYNALESRGVVDVLLILAHKYSHALPTLMCTLVGIIIVIGIYKMNVEFDKTSFGD  
NLNNTTQOTSMASLSSTNVSLEPVRSTPIASPOVSLPTARDRLNONGSGDN  
RLMSY"  
gene  
complement(16754, 19698)  
/gene="T23H2.4"  
complement(join(16754, 16999,17060, 17139,17676, 18125,  
18255, 18537,18591, 18727,18890, 18946,19003, 19047,  
19491, 19698))  
/gene="T23H2.4"  
/codon\_start=1  
/evidence-not\_experimental  
/protein\_id="AAC48198.1"  
/db\_xref="PDB:1703555"  
/db\_xref="GI:1703555"  
/translation="MTPFNAMLEKIIIVLVLDISYCELCFNEGTCEMDIRPNNVV  
SLAIPRHOHGLINISKTPDVTISLVVFLNMSNMKQSFGKMPFGILNTRFC  
TQGPLAOGKFAVSPSRVTSOKISKSDHMLIAYRQRMATRLILKMQTEIL  
LDTISSNEIERIKWHRSVIFPPIDVNOTSVIFGDNQOTMEDIIIVADEIDAPAN  
VMAQODSKFDEEDIRREIRRSKLSVCTPISCHRSSICSMQSTIOLLPKIV  
NEAGESTLTSVPRLPIDAPSFIDQMRNSQSVATSLSFNIGSNEVAVITKEDS  
SIPDSTQONGNRRVLSLESVKGIPVLLILKISITGANDFVAFSIDIADVDESKTL  
CDQVTPVQPOSNOVLRLALONKMLPESLILYOPITSTGPISTFPPPPSISFN  
SDERSSIASPGLIPISPTIVSARGIPFIPITSFNSNPLIPSNARLPPH"  
gene  
complement(20137, 21244)  
/gene="T23H2.5"  
complement(join(20137, 20223,20594, 20672,20750, 21062,  
21118, 21244))  
Query Match 22.5%; Score 34.6; DB 37; Length 23091;  
Best Local Similarity 64.2%; Pred. No. 1.5;  
Matches 52; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

AB026833  
LOCUS AB026833 3604 bp mRNA PRI 26-MAY-1999  
DEFINITION Home sapiens mRNA for chloride channel protein, complete cds.  
ACCESSION AB026833  
NID 94887600  
VERSION AB026833.1 GI:4887600  
KEYWORDS chloride channel protein.  
SOURCE Homo sapiens adult corneal epithelium cDNA to mRNA,  
clone\_lhb:lambda ZAP clone:lambda 13802.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Itoh, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.  
TITLE Isolation and characterization of chloride channel in human corneal  
epithelium  
JOURNAL Unpublished (1999)  
REFERENCE 2 (bases 1 to 3604)  
AUTHORS Itoh, R., Kawamoto, S. and Okubo, K.  
TITLE Direct Submision  
JOURNAL Submitted (28-Apr-1999) to the DDBJ/EMBL/GenBank databases. Relko  
Itoh, Institute for Molecular and Cellular Biology, Osaka  
University, Yamada-Oka 1-3, Suita, Osaka 565-0871, Japan  
(E-mail:relko@imcb.osaka-u.ac.jp, Tel:81-6-6879-7992,  
Fax:81-6-6877-1922)  
FEATURES  
source  
location/Qualifiers  
1..3604  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="lambda 13802"  
/clone\_lib="lambda ZAP"  
/dev\_stage="adult"  
/map="lp32"  
/tissue\_type="corneal epithelium"  
42..2873  
/codon\_start=1  
/product="chloride channel protein"  
/protein\_id="BA077810.1"  
/db\_xref="PDB:1041578"  
/db\_xref="PDB:94887601"  
/db\_xref="GI:4887601"  
/translation="MORSIAGPICNLKFTVLLVALSELPELAGVOLQDNGYGL  
TAINPVQENQNLINIKEMITEASEFYLPNATKRRVFRNKILPATMKANNSKIK  
DESTRKANVYIDWYGAHGDDEPTTLOYNGCKEGEYHFTFPLNDULIAGYSGRQ  
VFEHEMAHLMGVQDEYIDNRPEYINGQNIKVTSCSDIDIGFVCEGPGQENCII  
SLFKEGCTFIYNSQNTATASIMFQSSVVEFCNASTHNOEALONOMSLRSAM  
DVTISAGCFHSEFPNGTELPPEPFSLVQAGDKVCLVDVSSKMAADRLQDQA  
AEFYLMQVEIHTFVGIASFDSKGRIRAOIHOINNDKRLIVSLPTTVSAKKDTSI  
CSGLKGEVVEYERLAKGAYGSMLIYVSGDKLQNCILPYLVSSTIHSIALGSSAA  
PNLELSRLTGKLFKFPVDISNSMIDAFSKISSTGCDITQOH1QLSSTGVNVPFH  
OLKNTVYDNTVNDGNDTMELVQWAGSPEIILFDDGKRYTNNITNLTFRTASLWI  
PGTAPKGMVTLNNTNHSLOAKTVTSRASNSAVPAPYEAFAERDRLHPHVMV  
YANYVQGVPLINATVATVEPETDPTVLRLLDGGADADVAKNDGYSRYFSPAAV  
GRYSIKVHNSPSTSTPAHSIPGSHAVYGVGTNGNIONMAPRKSVGRNEEKRG  
FSRVSSGGSFVLTGYPACPHPDVPPKCIIDLEAKYVEBELTSLTAGEDDQOAT  
STEIMSRSLQNTQDFFNNALIVNTSKNPQOAGIREITFSPOJSTNGPEHPOGET  
HESRIRIYAIRAMDNRNSLQASVANSIAQAPLFPNPSDVPARDYDLILGVLITAMGLIG  
IICLIIVTHTLSRRKRDCKKENCFTKL"  
BASE COUNT 1113 a 740 c 732 g 1017 t 2 others  
ORIGIN  
Query Match 22.3%; Score 34.4; DB 9; Length 3604;  
Best Local Similarity 56.0%; Pred. No. 1.5;  
Matches 65; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Db 2686 CCCAGCGCCTCTGTTATTCCTCCCAATTCGATCTGTACCTCCAGAGATTAT 2741

## RESULT 11

LOCUS AF047838 3022 bp mRNA ROD 26-NOV-1998  
DEFINITION Mus musculus calcium-sensitive chloride conductance protein-1  
(mClCA1) mRNA, complete cds.

ACCESSION AF047838

NID 93925280

VERSION AF047838.1 GI:3925280

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

gene

CDS

Protein-1

Protein-1d

Protein-1e

Protein-1f

Protein-1g

Protein-1h

Protein-1i

Protein-1j

Protein-1k

Protein-1l

Protein-1m

Protein-1n

Protein-1o

Protein-1p

Protein-1q

Protein-1r

Protein-1s

Protein-1t

Protein-1u

Protein-1v

Protein-1w

Protein-1x

Protein-1y

Protein-1z

Protein-1aa

Protein-1ab

Protein-1ac

Protein-1ad

Protein-1ae

Protein-1af

Protein-1ag

Protein-1ah

Protein-1ai

Protein-1aj

Protein-1ak

Protein-1al

Protein-1am

Protein-1an

Protein-1ao

Protein-1ap

Db 2605 TCGACAGGCTGTGTTATTCCTCCACA 2635

## RESULT 12

LOCUS HSDJ399J4 15158 bp DNA HTG 03-JUN-1999  
DEFINITION Homo sapiens chromosome 6 clone DJ399J4, WORKING DRAFT SEQUENCE, in  
unoriented pieces.

ACCESSION HSDJ399J4

NID 95002654

VERSION AL078595.1 GI:5002654

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

White, S.

Submitted (03-JUN-1999) Wellcome Trust Genome Campus, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humangenesanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

IMPORTANT: This sequence is unfinished and does not necessarily

represent the correct sequence. Work on the sequence is in progress

and the release of this data is based on the understanding that the

sequence may change as work continues. The sequence may be

contaminated with foreign sequence from E.coli, yeast, vector,

phase etc. Order of segments is not known; 800 n's separate

segments. Unfinished: dj399J4 Contig\_ID: 00813 acc= length: 3481

bp Unfinished: dj399J4 Contig\_ID: 01546 acc= length: 1023 bp

Unfinished: dj399J4 Contig\_ID: 00188 acc= length: 1002 bp

Unfinished: dj399J4 Contig\_ID: 00584 acc= length: 1350 bp

Unfinished: dj399J4 Contig\_ID: 01002 acc= length: 22435 bp

Unfinished: dj399J4 Contig\_ID: 00990 acc= length: 9692 bp

Unfinished: dj399J4 Contig\_ID: 00994 acc= length: 1003 bp

Unfinished: dj399J4 Contig\_ID: 01814 acc= length: 32752 bp

Unfinished: dj399J4 Contig\_ID: 01256 acc= length: 1548 bp

Unfinished: dj399J4 Contig\_ID: 01494 acc= length: 1112 bp

Unfinished: dj399J4 Contig\_ID: 00690 acc= length: 1289 bp

Unfinished: dj399J4 Contig\_ID: 01198 acc= length: 1020 bp

Unfinished: dj399J4 Contig\_ID: 00316 acc= length: 8546 bp

Unfinished: dj399J4 Contig\_ID: 01924 acc= length: 1473 bp

Unfinished: dj399J4 Contig\_ID: 01130 acc= length: 1136 bp

Unfinished: dj399J4 Contig\_ID: 00404 acc= length: 48789 bp

Unfinished: dj399J4 Contig\_ID: 00404 acc= length: 1107 bp.

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

Location/Qualifiers

1. 15158

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="6"

/clone="DJ399J4"

BASE COUNT 44584 a 24859 c 24916 g 44386 t 12813 others

ORIGIN

Query Match

Best Local Similarity 51.8%; Pred. No. 8.1; Length 15158;

Matches 73; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Db 24906 GATCATATATCAATACCATGCAACACATCTATATATATCTCAATGATGATATCCCT 24965



QY 133 GCTCCTCTGCTAATATTCAT 153  
1 11 111 1111  
Db 24966 AATGATTGTGTAATAATCAT 24966

RESULT 13  
CEY37A1B/c 99083 bp DNA INV 18-DEC-1998  
LOCUS Caenorhabditis elegans cosmid Y37A1B, complete sequence.  
DEFINITION AL023835  
ACCESSION AL023835  
NID 93217862  
VERSION AL023835.1 GI:3217862  
KEYWORDS HTG.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 99083)  
REFERENCE  
AUTHORS McMurtry,A.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUN-1998) Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or rwenematode.wustl.edu  
2 (bases 1 to 99083)  
REFERENCE  
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,  
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,  
Coulson,A., Craxton,M., Dear,S., Du,Z., Dutilleul,R., Favello,A.,  
Fullon,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,  
Johnston,L., Jones,M., Kersey,J., Kirsten,J., Laister,N.,  
Latreille,P., Lightning,J., Lloyd,C., McMurtry,A., Mortimore,B.,  
O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra,A.,  
Saunders,D., Showkhen,R., Smalton,N., Smith,A., Sonnenhammer,E.,  
Staden,K., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,  
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,  
Wilkinson-Sproat,J. and Wohlman,P.  
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans  
JOURNAL Nature 368 (6466), 32-38 (1994)  
MEDLINE 94150718  
COMMENT Coding sequences below are predicted from computer analysis, using  
predictions from GeneFinder (P. Green, U. Washington), and other  
available information.  
For a graphical representation of this sequence and its analysis  
see:-  
http://webcace.sanger.ac.uk/cgi-  
bin/displaydb-wormcaceclass-sequence?object=Y37A1B  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring subclones.  
IMPORTANT: This sequence is not the entire insert of clone Y37A1B.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring subclones.  
The true left end of clone F52B11 is at 98980 in this sequence. The  
true right end of clone C27H2 is at 102 in this sequence. The start  
of this sequence (1..102) overlaps with the end of sequence 282256.  
The end of this sequence (98980..99083) overlaps with the start of  
sequence 282268.  
FEATURES  
source location/Qualifiers  
1..99083  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="IV"  
/clone="Y37A1B"  
gene complement(3796..9369)  
/gene="Y37A1B.14"  
CDS complement(join(3796..4266,4320..4779,5566..5827,

gene 5915..6003,7943..8023,8068..8260,8385..8682,8740..8909,  
8971..9075,9312..9369))  
/gene="Y37A1B.14"  
/note="predicted using GeneFinder; similar to Src homology  
domain 3 (2 domains)"  
/protein\_id="CA19485.1"  
/codon\_start=1  
/db\_xref="PID:e1350268"  
/db\_xref="PID:g3880771"  
/db\_xref="GI:3880771"  
/translation="MTPRKNGCTGMEPRKTKTOYYSOCQFCTLKIMSETEVLAGNV  
ICNCLAEKSAVOQDPKRYMTKLSKTPSRKWTLSNNILMLILLLSALFEISQGF  
SMALNEPQDKNPQNTIESYRSFYVNDSPSISKENRRLALRLEQMLHPDEARPLY  
IVLAGLNNEFSEKLNKAIQNSKSETNSNNPIEIGNOLEGERLHALLNNALASQMT  
RTAVLTNIDLAMDVAVLHAFADHGTPEIPKTVLLAVSEETYSQNSCDGKILKL  
VTNRWIEGNGYNDIPPIVARIISYIOPEQVYCSIEAYTSIRSDLSFAFSIFVAK  
PSRNPGLDGFVNDNAVOIGRGLPEALHVLNLYSAEISPMKSTPTDFSEADVFR  
RTPATAPAASTNAVSPIYARAAYVDPOGEFENELISFADELISRRRTADAMEGSLG  
SARVGIPPTSFVQIIVLPPDDTSLQNNRSTTTENDIGEFANVRHAFSGRODEL  
VNAGDITVLRNVDENWCKDPDTEKGLVPGFLEVYLEEDDDNQSNAMRGIG  
GTSDFTEGSSNNRSSFISNAESSDWRPACRINTDVAEFAVPAAPASPAQNMAT  
FGEEMVAPVQIAKTSAPARPPPKQSSIOSPIINDIGNVADMGMTDTPKPSNPAIQSA  
GSDEIIGISATGASNDEDKRCRIIEELISSEIQFISDINSYEV"  
complement(9530..11012)  
/gene="Y37A1B.13"  
complement(join(9530..10217,10268..10451,10646..11012))  
/gene="Y37A1B.13"  
/note="predicted using GeneFinder"  
/codon\_start=1  
/protein\_id="CA19484.1"  
/db\_xref="PID:e1350267"  
/db\_xref="PID:g3880770"  
/db\_xref="GI:3880770"  
/translation="MKKFAEKMFLLKRFYVQCFIFKFRYQCINLFEFGVISAGYFDV  
SKNQITSDICFSISFSTSHLSNILEYSEKQGFYIIFVILLNQLVVDHSLSP  
MELKCLPYTCGGEDIFNYHALYDFPNKIEIGHALMESVYSIKSHWHNHSQKPLY  
LSRFGGTGCKNVTETIIVNNTYSSGHSPPVNFVANNPNKTYIEDYLETEDOL  
IRARRCORSIFITDETDKLOSELQVAKPFLUDYPAFGVDPRKTIIFILSKNGSE  
IANTALHHEHGKIRSOLELKHFERITLMSFENEGGIRNTDMSNLDIDHFIPLPL  
SKFVSQICQVHLKRRGRHDLAKDGEFMQVRVLDLLEFPESKIFSSGCKRVNAKD  
LEIKMFSLSKREFNDEL"  
complement(11360..14313)  
/gene="Y37A1B.12"  
complement(join(11360..12047,12101..12284,13190..13286,  
14275..14313))  
/gene="Y37A1B.12"  
/codon\_start=1  
/protein\_id="CA19495.1"  
/db\_xref="PID:e1350278"  
/db\_xref="PID:g3880781"  
/db\_xref="GI:3880781"  
/translation="MLVIRNFPPLILINAKNTRVTKNRTNMNKLNNMNLFFELSTENG  
LELRDLSRFYGGHVLVDVNRVIRKSHMHNEHQKPLSFGSGPTGKRYVEILA  
KNTFRSGISPYVYFVATKDFPNKRIEDYKILKCOLIOSAGCGRSIFVPEVVK  
LQSLNQTIPDFDFYPAVEVDPRKTIIFILSKGSSSEIANTLHERRNLKRSOLE  
LKHRETLMSHAFNEKGLRTEILSOLIDHTIPLPLSFYSQICQVHLKRRGRH  
DLANDGEFMQVRVLDLLEFPESKVFSSGCKRVADATELEISXMGFSLNSEKEFND  
L"  
30545..38872  
/gene="Y37A1B.11"  
join(30545..30601,30704..30857,31351..31444,31942..32011,  
32903..33131,33890..34048,34413..34718,36736..36894,  
37068..37261,37636..37806,38535..38872)  
/gene="Y37A1B.11"  
/note="similar to potassium channel protein"  
/codon\_start=1  
/protein\_id="CA19494.1"  
/db\_xref="PID:e1350277"  
/db\_xref="PID:g3880780"  
/db\_xref="GI:3880780"  
/translation="MKLLTKSVFSRDKHLQKATPLPVHFLMIVSGAAYRIGALVY  
RSLSEPTVSTSEKKTIVRRHRRVLTNQHPTPTLLEQRHRRRRHRETALEHLISK  
LSREKRAAHIMRSKCVISVIKMSMECSFDLDELKLVKALDECHHVAENHTVHN





Db 7837 ACCGTGTATCAATAGCCTTATTTATTCCTTACTACATCATCAAGCCTAGACT 7896

Db 2658 CACATTGTACAGCAATCAATTTATTCCTTAC 2691

Search completed: August 5, 1999, 23:14:56  
Job time: 7779 sec

RESULT 15  
AF001261 3317 bp mRNA MAM 19-NOV-1997  
LOCUS Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA,  
DEFINITION complete cds.  
ACCESSION AF001261  
VERSION 92623762  
KEYWORDS AF001261.1 GI:2623762  
SOURCE Bos taurus.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 3317)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
TITLE Goodwin, A. and Pauli, B.U.  
Cloning and characterization of Lu-ECAM-1 suggest it is an  
endothelial chloride channel  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3317)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
TITLE Goodwin, A. and Pauli, B.U.  
Direct Submission  
Submitted (25-APR-1997) Vet. Pathology, Cornell U., WMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES  
source Location/Qualifiers  
1.3317  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="aortic endothelial cells"  
/clone="1"  
/tissue\_type="lung"  
/note="endothelial adhesion molecule; chloride channel"  
CDS  
/product="lu-ECAM-1"  
/protein\_id="AAB86529.1"  
/db\_xref="PID:92623763"  
/db\_xref="GI:2623763"  
/translation="MVLCLNVILFLTHLHPGKSSMVLINNGDGIYIAINPSVE  
DEKLEIKEMTEASTYLFHATKRRVYFRNVSILPMTWKSSEFFIKPQSTIDAD  
VIVANPLYKGDDEPTLOYGRGEGKGIHFTPNFLINNFHLYGSRGVFVEMAH  
RGIPEYVNDQDPFYISRWKTIETATRCSTHITGINVFKKPGSCITSLCRDSQTG  
LTERKCTFLPKKSQTAKESITMEMPISHSVTECTEKTHTNEAPNLQNMKNGKSTWD  
IMNSVDONTSPMTENPPTHPFSLKSKQRYVCLYDKSGMSAEDRLFQMNQAE  
LYLQVLEKSLVGMVFDVSAEIQNLHTRITDNYOKITAKLPYANGGTSGRGL  
KAGFOALIHSDQSTSGSEIILLDGEDNEINSCFEDVKRSGALITHITAGPSAKETE  
TLSMTGIRFFRANKDITGLTNAFSRISRSISGSIIOQAIOLESKALITGRKRVNGTV  
PVSTVGNDFEYVVTITOKPEIYLOPKKGYKTSDFEKDKLINSARLQIPGIAET  
GTVTYSILNNHASOMLTAVTTTRARSPITPIVATAHMSOHTAHYPSPIVAYQVSO  
GFLDVLGISTYIAIETREDGCHVTELEMDNGARDYKNDGIYRFTDYDNGRYSGLK  
VHAQRNNITARLNLROPONKVLVYIPGYENGKTIILNDPREVVDOLAKAKIEDSRIT  
SGSGFTVSGAPPGNHPSPKSTIDLEAKFKEDYIQLSMTAPGNVLDKGRANSYII  
RISKSEMDROEDENATLVNTSLIPREAGSKSEFEKPEHFRVENGKTFYISVQAIN  
EANLISESHVQAIKFIPLPEDVDHDLGTISBITLAIIGLPMIFSVF"

BASE COUNT 1084 a 676 c 634 g 923 t  
ORIGIN

Query Match 20.5%; Score 31.6; DB 3; Length 3317;  
Best Local Similarity 58.5%; Pred. No. 10;  
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 1 ACAGATCTTTTCAATTCAGGCTGTGATAGTCGATCGAATCAGAAATATCC 60  
DB 2598 ACCAAATTCATATTCAGTCAAGCCATCAAGCAATTCATCTCAGAGGTTCT 2657  
OY 6A ACATTGACAGAGTANCTTGTATTCTCCAC 94

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:59 ; Search time 3059.95 Seconds  
(without alignments)  
99.273 Million cell updates/sec

Title: US-09-049-696-15

Perfect score: 154  
Sequence: 1 ACAAGTCTTTCATTGCTAT.....TCCTTGCCATATATTCATA 154

Scoring table: OLIGO\_NWC

Searched: 2546578 seqs, 986266752 residues

Database : EST.\*

1: em\_est1:.\*  
2: em\_est2:.\*  
3: em\_est3:.\*  
4: em\_est4:.\*  
5: em\_est5:.\*  
6: em\_est6:.\*  
7: em\_est7:.\*  
8: em\_est8:.\*  
9: em\_est9:.\*  
10: em\_est10:.\*  
11: em\_est11:.\*  
12: em\_est12:.\*  
13: em\_est13:.\*  
14: em\_est14:.\*  
15: em\_est15:.\*  
16: em\_est16:.\*  
17: em\_est17:.\*  
18: em\_est18:.\*  
19: em\_est19:.\*  
20: gb\_est1:.\*  
21: gb\_est2:.\*  
22: gb\_est3:.\*  
23: gb\_est4:.\*  
24: gb\_est5:.\*  
25: gb\_est6:.\*  
26: gb\_est7:.\*  
27: gb\_est8:.\*  
28: gb\_est9:.\*  
29: gb\_est10:.\*  
30: gb\_est11:.\*  
31: gb\_est12:.\*  
32: gb\_est13:.\*  
33: gb\_est14:.\*  
34: gb\_est15:.\*  
35: gb\_est16:.\*  
36: gb\_est17:.\*  
37: gb\_est18:.\*  
38: gb\_est19:.\*  
39: gb\_est20:.\*  
40: gb\_est21:.\*  
41: gb\_est22:.\*  
42: gb\_est23:.\*  
43: gb\_est24:.\*  
44: gb\_est25:.\*  
45: gb\_est26:.\*  
46: gb\_est27:.\*  
47: gb\_est28:.\*  
48: gb\_est29:.\*  
49: gb\_est30:.\*  
50: gb\_est31:.\*  
51: gb\_est32:.\*  
52: em\_est20:.\*  
53: em\_est21:.\*

54: em\_est22:.\*  
55: em\_est23:.\*  
56: em\_est24:.\*  
57: em\_est25:.\*  
58: em\_est26:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	154	100.0	378 31	AA296955
2	154	100.0	490 34	AA508854
3	154	100.0	716 49	A1660234
4	154	100.0	734 49	A1660957
5	154	100.0	335 51	A1721121
6	154	100.0	501 51	A1721275
7	67	43.5	255 20	D25727
8	57	37.0	392 50	A1687981
9	54	35.1	310 31	AA297150
10	20	13.0	161 48	A1382072
11	19	12.3	435 50	A0065204
12	17	11.0	205 20	F01472
13	17	11.0	334 20	Z39640
14	17	11.0	299 21	T73014
15	17	11.0	504 22	R51332
16	17	11.0	357 24	H85938
17	17	11.0	589 25	N95399
18	17	11.0	438 26	W85901
19	17	11.0	386 28	AA085271
20	17	11.0	127 29	AA153556
21	17	11.0	424 29	AA180481
22	17	11.0	634 29	AA191695
23	17	11.0	390 30	AA255524
24	17	11.0	377 31	AA312640
25	17	11.0	276 32	AA368412
26	17	11.0	296 33	AA405228
27	17	11.0	313 33	AA406127
28	17	11.0	318 33	AA415963
29	17	11.0	362 35	AA535346
30	17	11.0	490 39	AA882303
31	17	11.0	408 40	AA985645
32	17	11.0	645 44	AU038165
33	17	11.0	403 47	A1494399
34	17	11.0	421 48	A1594699
35	17	11.0	127 48	A1603935
36	17	11.0	476 49	A1650501
37	17	11.0	74 49	A1661986
38	17	11.0	660 53	HSM006004
39	16	10.4	360 23	D60808
40	16	10.4	445 23	H58997
41	16	10.4	360 24	D67941
42	16	10.4	421 24	N29457
43	16	10.4	486 25	W14080
44	16	10.4	483 28	AA058693
45	16	10.4	291 28	AA058787

## ALIGNMENTS

RESULT 1  
AA296955 378 bp mRNA  
LOCUS EST112726 Colon I Homo sapiens CDNA 5' end, mRNA sequence.  
DEFINITION AA296955  
ACCESSION G1949515  
NID  
VERSION AA296955.1 GI:1949515  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 378)  
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodet,A., Guelm,C.L., Hanna,M.C., Hedblom,E., Hinkle,F.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Benatarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dunke,D., Feng,D.-F., Fertle,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M., and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
96026280  
On Sep 12, 1996 this sequence version replaced gi:1288227.  
Other ESTs: THC167738  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genome Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tldb/hgi/hgi.html>)  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..378  
/organism="Homo sapiens"  
/db\_xref="ATCC (host):127834"  
/db\_xref="taxon:9606"  
/clone\_lib="Colon I"  
/dev\_stage="adult"  
/note="Organ: colon; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 112 a 79 c 64 g 119 t 4 others  
ORIGIN

Query Match 100.0%; Score 154; DB 31; Length 378;  
Best Local Similarity 100.0%; Pred. No. 3.4e-77;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTCATTCGATTCAGGCTGTTGATAGGTCGATCGAATCAAAATATCC 60  
|||||  
DB 94 ACAGATCTTTTCATTCGATTCAGGCTGTTGATAGGTCGATCGAATCAAAATATCC 153  
QY 61 AACATTGCACAGATATCTTTGTTTATTCCTCCACAGACTCCGCCAGAGACCTAGTCT 120  
|||||  
DB 154 AACATTGCACAGATATCTTTGTTTATTCCTCCACAGACTCCGCCAGAGACCTAGTCT 213  
QY 121 GATGAACGCTGCTCTGCTCTGCTATATTCATA 154  
|||||  
DB 214 GATGAACGCTGCTCTGCTCTGCTATATTCATA 247

RESULT 2  
AA508854/c 490 bp mRNA EST 18-AUG-1997  
LOCUS AA508854.1  
DEFINITION n12m10.s1 NCI\_CGAP\_Co4 Homo sapiens cDNA clone IMAGE:968803 3',

ACCESSION mRNA sequence.  
NID AA508854  
VERSION 52246357  
KEYWORDS AA508854.1 GI:2246357  
SOURCE EST  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 490)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1397407.  
JOURNAL COMMENT  
TITLE  
AUTHORS  
REFERENCE  
TITLE  
AUTHORS  
JOURNAL  
COMMENT  
On Sep 12, 1996 this sequence version replaced gi:1397407.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: [www.bio.lnl.gov/dbtrp/image/image.html](http://www.bio.lnl.gov/dbtrp/image/image.html)  
Insert Length: 856 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 404.  
Location/Qualifiers  
1..490  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:968803"  
/clone\_lib="NCI\_CGAP\_Co4"  
/sex="pooled"  
/tissue\_type="colon"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled colon tumor tissue, and was then primed with a Not I - Oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. This library is not normalized. Library constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 155 a 82 c 89 g 164 t  
ORIGIN

Query Match 100.0%; Score 154; DB 34; Length 490;  
Best Local Similarity 100.0%; Pred. No. 3.2e-77;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTCATTCGATTCAGGCTGTTGATAGGTCGATCGAATCAAAATATCC 60  
|||||  
DB 488 ACAGATCTTTTCATTCGATTCAGGCTGTTGATAGGTCGATCGAATCAAAATATCC 429  
QY 61 AACATTGCACAGATATCTTTGTTTATTCCTCCACAGACTCCGCCAGAGACCTAGTCT 120  
|||||  
DB 428 AACATTGCACAGATATCTTTGTTTATTCCTCCACAGACTCCGCCAGAGACCTAGTCT 369  
QY 121 GATGAACGCTGCTCTGCTCTGCTATATTCATA 154  
|||||  
DB 368 GATGAACGCTGCTCTGCTCTGCTATATTCATA 335

RESULT 3  
A1660234/c 716 bp mRNA EST 10-MAY-1999  
LOCUS A1660234.1  
DEFINITION web8902.x1 Soares\_Dieckgraefe\_colon\_NHCD Homo sapiens cDNA clone

IMAGE:2346290 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA  
sequence.  
ACCESSION A1660234  
NID 94763804  
VERSION A1660234.1 GI:4763804  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 716)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949219.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 387.

FEATURES  
Source  
1..716  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2346290"  
/clone\_lib="Soares\_Dieckgraefe\_colon\_NHNC"  
/tissue\_type="colonic mucosa from 3 patients with Crohn's  
disease"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pRT3D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCATCTGAGTGGAGCGGCCGCCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pRT3D vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieck@im.wustl.edu); colonic mucosa represents a range of  
disease involvement from moderate to severe Crohn's  
disease; samples include both perforating (fistulas) and  
non-perforating samples. Library constructed by Bento  
Soares and M. Fatima Bonaldo.

BASE COUNT 220 a 123 c 140 g 233 t  
ORIGIN

Query Match 100.0%; Score 154; DB 49; Length 716;  
Best Local Similarity 100.0%; Pred. No. 2.9e-77;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTCAATGCTATTCAGGCTGTGATAGGTCGATGAAATCAGAAATATCC 60  
|||||  
DB 485 ACAGATCTTTTCAATGCTATTCAGGCTGTGATAGGTCGATGAAATCAGAAATATCC 426  
|||||

QY 61 AACATTGCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTATGCTCT 120  
|||||  
DB 425 AACATTGCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTATGCTCT 366  
|||||

QY 121 GATGAACGCTGCTGCTTGTCTCTATATTCATA 154  
|||||  
DB 365 GATGAACGCTGCTGCTTGTCTCTATATTCATA 332  
|||||

RESULT 4  
A1660957/c 734 bp mRNA EST 10-MAY-1999  
LOCUS A1660957  
DEFINITION AF20408.x1 Soares\_Dieckgraefe\_colon\_NHNC Homo sapiens cDNA clone  
IMAGE:2351151 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA

sequence.  
ACCESSION A1660957  
NID 94764540  
VERSION A1660957.1 GI:4764540  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 734)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980033.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 477.

FEATURES  
Source  
1..734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="X"  
/clone="IMAGE:2351151"  
/clone\_lib="Soares\_Dieckgraefe\_colon\_NHNC"  
/tissue\_type="colonic mucosa from 5 ulcerative colitis  
patients"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pRT3D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCATCTGAGTGGAGCGGCCGCCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pRT3D vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieck@im.wustl.edu); colonic mucosa represents a range of  
disease involvement from mild cryptitis to severe  
ulceration, fibrosis, and degeneration. Library  
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 200 a 143 c 159 g 228 t  
ORIGIN

Query Match 100.0%; Score 154; DB 49; Length 734;  
Best Local Similarity 100.0%; Pred. No. 2.9e-77;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTCAATGCTATTCAGGCTGTGATAGGTCGATGAAATCAGAAATATCC 60  
|||||  
DB 281 ACAGATCTTTTCAATGCTATTCAGGCTGTGATAGGTCGATGAAATCAGAAATATCC 222  
|||||

QY 61 AACATTGCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTATGCTCT 120  
|||||  
DB 221 AACATTGCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTATGCTCT 162  
|||||

QY 121 GATGAACGCTGCTGCTTGTCTCTATATTCATA 154  
|||||  
DB 161 GATGAACGCTGCTGCTTGTCTCTATATTCATA 128  
|||||

RESULT 5  
A1721121 335 bp mRNA EST 10-JUN-1999  
LOCUS A1721121/c  
DEFINITION as73d08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
IMAGE:2334351 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA  
sequence.





```

ACCESSION   D25727
VERSION     9500422
KEYWORDS    D25727.1 GI:500422
SOURCE      EST.
ORGANISM    human.
             Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 255)
AUTHORS     Okubo, K., Yoshii, J., Yokouchi, H., Kameyama, M. and Matsubara, K.
TITLE       Global analysis of gene expression in colon mucosa: a large scale
             random cDNA sequencing analysis
JOURNAL     Unpublished (1994)
COMMENT     Contact: Okubo, K., Itoh, K., Yoshii, J., Yokouchi, H. and Matsubara, K.
             Institute for Molecular and Cellular Biology
             Osaka University
             3-1 Yamadaoka, Suita, Osaka 565, Japan.
FEATURES    source
             location/Qualifiers
             1..255
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="cm2037"
             /clone.lib="Human colon mucosa"
             /note="Adult male, tissue_type = colon mucosa"
BASE COUNT  76 a 52 c 39 g 80 t 8 others
ORIGIN
Query Match 43.5%; Score 67; DB 20; Length 255;
Best Local Similarity 100.0%; Pred. No. 6.7e-28;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 88 CCTCCAGACAGCTCCGACAGACACCTACTGATGATGAAGCTGCTCTGCTCTAAT 147
    |||||||
DB 49 CCTCCAGACAGCTCCGACAGACACCTACTGATGATGAAGCTGCTCTGCTCTAAT 108
OY 148 ATTCATA 154
    |||||||
DB 109 ATTCATA 115
RESULT 8
A1687981 392 bp mRNA EST 27-MAY-1999
LOCUS wa77f06.x1 Soares.NFL.T.GBC.S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2302211 3' similar to TR:088826 088826 GOB-5 PROTEIN. ;, mRNA
sequence.
ACCESSION A1687981
NID 94899275
VERSION A1687981.1 GI:4899275
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 392)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2947973.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40UP from gibco.
Location/Qualifiers
1..392
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="IMAGE:2302211"
/clone.lib="Soares.NFL.T.GBC.S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH19W, testis NHT, and B-cell
NCL-CGAP-GCBI) were mixed, and ss cicles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
T.M.A.G.E. clones 297480-302087 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonalido."
BASE COUNT 125 a 69 c 67 g 131 t
ORIGIN
Query Match 37.0%; Score 57; DB 50; Length 392;
Best Local Similarity 100.0%; Pred. No. 2.8e-22;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 98 CTCGCCAGAGACACTAGTCTGATGAAGCTGCTGCTCTGCTATATTCATA 154
    |||||||
DB 355 CTCGCCAGAGACACTAGTCTGATGAAGCTGCTGCTCTGCTATATTCATA 299
RESULT 9
AA297150 310 bp mRNA EST 18-APR-1997
LOCUS EST112734 Colon I Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION AA297150
ACCESSION 91949524
NID AA297150.1 GI:1949524
VERSION AA297150.1 GI:1949524
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 310)
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulmer, R.A.,
             Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
             White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-vel, C.,
             Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,
             Fitzgerald, L.M., Fitchugh, W.M., Fritchman, J.L., Geoghegan, N.S.,
             Glodex, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
             Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
             Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
             Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
             Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
             Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
             Dinke, D., Feng, D.-F., Ferlie, A., Fischer, C., Hastings, G.A.,
             He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,
             Korzak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H.,
             Raymond, L., Wei, T.F., Wing, J.J., Xu, C., Yu, G.L., Ruen, S.M.,
             Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
             Fraser, C.M., and Venter, J.C.
TITLE Initial assessment of human gene diversity and expression patterns
             based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT On Sep 12, 1996 this sequence version replaced gi:1288566.
Other ESTs: TRC167738
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene

```

Index (<http://www.figr.org/tcb/hgl/hgl.html>)  
Seq primer: M3 Reverse.

FEATURES  
Source  
Location/Qualifiers

1..310  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):127942"  
/db\_xref="taxon:9606"  
/clone\_1lb="Colon 1"  
/dev\_stage="adult"  
/note="Organ: colon; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 97 a 61 c 58 g 91 t 3 others  
ORIGIN

Query Match 35.1%; Score 54; DB 31; Length 310;  
Best Local Similarity 100.0%; Pred. No. 1.5e-20;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACAGATCTTTCATTCGATTCAGCGCTGTGATAGTCGATCGAATCAGAA 54  
|||||  
DB 231 ACAGATCTTTCATTCGATTCAGCGCTGTGATAGTCGATCGAATCAGAA 284

RESULT 10  
A1582072/c 161 bp mRNA EST 06-APR-1999  
LOCUS at96a11.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
DEFINITION IMAGE:2173436 3', mRNA sequence.  
ACCESSION A1582072  
NID 94567969  
VERSION A1582072.1 GI:4567969  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 161)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S.,  
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F.,  
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
Washington NCI Human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Mar 19, 1997 this sequence version replaced gi:1900677.

TITLE  
JOURNAL  
COMMENT

FEATURES  
Source  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.jnl.gov](mailto:info@image.jnl.gov)) for further information.  
Seq primer: 40UP from Gibco  
High quality sequence stop: 115.  
Location/Qualifiers

1..161  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2173436"  
/clone\_1lb="Barstead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: pRT3D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTCGATCTTCGATTCGATTCAGCGCGCCCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[5' ATCTACTAGTAT 3' and 5' ATCTAGAG 3'], digested  
with Not I and cloned into the Not I and Eco RI sites of  
the modified pRT3D vector. Library constructed by Bob

BASE COUNT 55 a 24 c 29 g 53 t  
ORIGIN

Query Match 13.0%; Score 20; DB 48; Length 161;  
Best Local Similarity 100.0%; Pred. No. 0.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 135 TCCCTGCTCAATATTCATA 154  
|||||  
DB 161 TCCCTGCTCAATATTCATA 142

RESULT 11  
A0065204/c 435 bp mRNA EST 20-MAY-1999  
LOCUS A0065204 Rice cDNA from immature leaf including apical meristem  
DEFINITION Oryza sativa cDNA clone E60611\_1A, mRNA sequence.  
ACCESSION A0065204  
NID 9480905  
VERSION A0065204.1 GI:4880905  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
REFERENCE 1 (bases 1 to 435)  
Sasaki, T. and Yamamoto, K.  
Rice cDNA from immature leaf including apical meristem  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187111.

CONTACT: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program  
2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abrr.affrc.go.jp  
PROJECT "RGP"  
Location/Qualifiers

FEATURES  
Source  
1..435  
/organism="Oryza sativa"  
/strain="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="E60611\_1A"  
/clone\_1lb="Rice cDNA from immature leaf including apical  
meristem"  
/dev\_stage="Immature"  
/note="Organ: leaf; Immature leaf including apical  
meristem (under long day condition)"  
BASE COUNT 115 a 102 c 110 g 107 t 1 others  
ORIGIN

Query Match 12.3%; Score 19; DB 50; Length 435;  
Best Local Similarity 100.0%; Pred. No. 0.89;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 TCTTGTTCCTCCAC 94  
|||||  
DB 393 TCTTGTTCCTCCAC 375

RESULT 12  
F01472 205 bp mRNA EST 01-FEB-1995  
LOCUS F01472  
DEFINITION HSC04B022 normalized infant brain cDNA Homo sapiens cDNA clone  
c-04b02 3', mRNA sequence.

ACCESSION F01472  
NID 9645029  
VERSION F01472.1 GI:645029  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes,M.D., Duprat,S., Houligatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouillot,Y., Sebastiani-Kabatchis,C. and Tessier,A.  
TITLE IMAGE: molecular integration of the analysis of the human genome and its expression  
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
MEDLINE 95277534  
COMMENT On Oct 24, 1994 this sequence version replaced.

FEATURES  
SOURCE  
Contact: Genethon  
Genethon Centre de recherche sur le Genome Humain  
1, rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read. removed at sequence 5'end  
Genexpress\_library\_id: C; Genexpress\_sequence\_id: alc-04b02  
Seq primer: (-21)M13-universal.  
Location/Qualifiers  
1..205  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="c-04b02"  
/clone\_lib="normalized infant brain cDNA"  
/sex="Female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lafmid BA; Site:1: HindIII; Site:2: NotI; sex=Female; dev\_stage=3 months old; Isolate-muscular atrophy patient; tissue\_type=total brain; total mRNA was oligo-(GT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Saeres, Psychiatry Dept. Columbia University, USA. Normalization\_method: Bento Soares, P.N.A.S in press"

BASE COUNT 52 a 40 c 56 g 57 t  
ORIGIN

Query Match 11.0%; Score 17; DB 20; Length 205;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 51 AGAATATCCACATG 67  
|||||  
Db 31 AGAATATCCACATG 47

RESULT 13  
LOCUS 239640 334 bp mRNA EST 07-NOV-1994  
DEFINITION HSC1FD112 normalized infant brain cDNA Homo sapiens cDNA clone c-1fd11 3, mRNA sequence.  
ACCESSION 239640  
NID 9565435  
VERSION 239640.1 GI:565435  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
JOURNAL 1 (bases 1 to 334)

AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes,M.D., Duprat,S., Houligatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouillot,Y., Sebastiani-Kabatchis,C. and Tessier,A.  
TITLE IMAGE: molecular integration of the analysis of the human genome and its expression  
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
MEDLINE 95277534  
COMMENT

FEATURES  
SOURCE  
Contact: Genethon  
Genethon Centre de recherche sur le Genome Humain  
1, rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read. 18 T removed at sequence 5'end  
Genexpress\_library\_id: C; Genexpress\_sequence\_id: alc-1fd11  
Seq primer: (-21)M13-universal.  
Location/Qualifiers  
1..334  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="12: 22"  
/clone="c-1fd11"  
/clone\_lib="normalized infant brain cDNA"  
/sex="Female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lafmid BA; Site:1: HindIII; Site:2: NotI; sex=Female; dev\_stage=3 months old; Isolate-muscular atrophy patient; tissue\_type=total brain; total mRNA was oligo-(GT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Saeres, Psychiatry Dept. Columbia University, USA. Normalization\_method: Bento Soares, P.N.A.S in press"

BASE COUNT 91 a 66 c 91 g 85 t 1 others  
ORIGIN

Query Match 11.0%; Score 17; DB 20; Length 334;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 51 AGAATATCCACATG 67  
|||||  
Db 31 AGAATATCCACATG 47

RESULT 14  
LOCUS T73014 299 bp mRNA EST 02-MAR-1995  
DEFINITION yc69d04.s1 Stratagene liver (#937224) Homo sapiens cDNA clone IMAGE:85927 3', mRNA sequence.  
ACCESSION T73014  
NID 9689689  
VERSION T73014.1 GI:689689  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapelli,B., Chisoe,S., Dietrich,N., Dubuque,T., Favelli,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Roeding,T., Schellenberg,K., Soares,M.B., Tan,F., Thierly-Meg,J., Trevaalis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.  
TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)

MEDLINE  
97044478

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Insert Size: 2026  
High quality sequence stops: 241 Source: IMAGE Consortium, LNL This  
clone is available royalty-free through LNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.  
Insert Length: 2026 Std Error: 0.00  
Seq primer: -21ml3

High quality sequence stop: 241.  
Location/Qualifiers

FEATURES  
source

```
1. .299
/organism="Homo sapiens"
/db_xref="GDB:502984"
/cb_xref="taxon:9606"
/clone="IMAGE:85927"
/clone_lib="Stratagene liver (#937224)"
/sex="male"
/dev_stage="49 years old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: liver; Vector: pBluescript SK; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Hepatotomy from normal male caucasian. Average
insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTCAGCTTTTCTTTTCTTTT 3'"
79 a 60 c 76 g 81 t 3 others
```

BASE COUNT  
ORIGIN

Query Match 11.0%; Score 17; DB 21; Length 299;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 51 AGAATATCCACATTG 67  
|||||  
Db 18 AGAATATCCACATTG 34

RESULT 15

R51322 504 bp mRNA EST 18-MAY-1995  
LOCUS R51322  
DEFINITION Y72707.81 Soares infant brain INIB Homo sapiens cDNA clone  
IMAGE:38685 3', mRNA sequence.

ACCESSION R51322  
NID 9813224  
VERSION R51322.1 GI:813224  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 504)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1847

High quality sequence stops: 385 Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1847 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 385.  
Location/Qualifiers

FEATURES  
source

```
1. .504
/organism="Homo sapiens"
/db_xref="GDB:411226"
/cb_xref="taxon:9606"
/clone="IMAGE:38685"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: latmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dt) primer [5';
AAGTGAAGATTCGGCGCCGAGAAATTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT  
ORIGIN

Query Match 11.0%; Score 17; DB 22; Length 504;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 51 AGAATATCCACATTG 67  
|||||  
Db 44 AGAATATCCACATTG 60

Search completed: August 6, 1999, 09:23:03  
Job time: 6254 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 13:33:25 ; Search time 3075.15 Seconds  
(without alignments)  
98.782 Million cell updates/sec

Title: US-09-049-696-15

Perfect score: 154  
Sequence: 1 ACAGATCTTTCATGCTAT.....TCCTGCTCATATATTCATA 154

Scoring table: IDENTITY\_MUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:\*

1: em\_est1:.\*  
2: em\_est2:.\*  
3: em\_est3:.\*  
4: em\_est4:.\*  
5: em\_est5:.\*  
6: em\_est6:.\*  
7: em\_est7:.\*  
8: em\_est8:.\*  
9: em\_est9:.\*  
10: em\_est10:.\*  
11: em\_est11:.\*  
12: em\_est12:.\*  
13: em\_est13:.\*  
14: em\_est14:.\*  
15: em\_est15:.\*  
16: em\_est16:.\*  
17: em\_est17:.\*  
18: em\_est18:.\*  
19: em\_est19:.\*  
20: gb\_est1:.\*  
21: gb\_est2:.\*  
22: gb\_est3:.\*  
23: gb\_est4:.\*  
24: gb\_est5:.\*  
25: gb\_est6:.\*  
26: gb\_est7:.\*  
27: gb\_est8:.\*  
28: gb\_est9:.\*  
29: gb\_est10:.\*  
30: gb\_est11:.\*  
31: gb\_est12:.\*  
32: gb\_est13:.\*  
33: gb\_est14:.\*  
34: gb\_est15:.\*  
35: gb\_est16:.\*  
36: gb\_est17:.\*  
37: gb\_est18:.\*  
38: gb\_est19:.\*  
39: gb\_est20:.\*  
40: gb\_est21:.\*  
41: gb\_est22:.\*  
42: gb\_est23:.\*  
43: gb\_est24:.\*  
44: gb\_est25:.\*  
45: gb\_est26:.\*  
46: gb\_est27:.\*  
47: gb\_est28:.\*  
48: gb\_est29:.\*  
49: gb\_est30:.\*  
50: gb\_est31:.\*  
51: gb\_est32:.\*  
52: em\_est20:.\*  
53: em\_est21:.\*

54: em\_est22:.\*  
55: em\_est23:.\*  
56: em\_est24:.\*  
57: em\_est25:.\*  
58: em\_est26:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	154	100.0	378	31	AA296955	AA296955 EST112726
2	154	100.0	490	34	AA508854	AA508854 n122h10.s
3	154	100.0	716	49	A1660234	A1660234 we68q02.x
4	154	100.0	734	49	A1660957	A1660957 wf20d08.x
5	154	100.0	335	51	A1721121	A1721121 as73d08.x
6	154	100.0	501	51	A1721275	A1721275 as82h08.x
7	112	72.7	255	20	D25727	D25727 HUMG504094
8	90.8	59.0	392	50	A1687981	A1687981 wa77f06.x
9	75.2	48.8	175	47	A1506262	A1506262 vq50e03.x
10	73.6	47.8	431	30	AA238284	AA238284 my34g09.r
11	73.6	47.8	418	35	AA596289	AA596289 vo26a08.r
12	73.6	47.8	552	37	AA688953	AA688953 vs04h07.r
13	73.6	47.8	301	37	AA689856	AA689856 vt62c11.r
14	73.6	47.8	646	37	AA691335	AA691335 vs14h06.r
15	73.6	47.8	421	37	AA691586	AA691586 vs12g08.r
16	73.6	47.8	482	37	AA711228	AA711228 vt70c11.r
17	73.6	47.8	488	37	AA733415	AA733415 vt75b06.r
18	73.6	47.8	368	37	AA734161	AA734161 vs19g06.r
19	73.6	47.8	297	49	AA839323	AA839323 v040f07.r
20	73.6	47.8	297	49	A1641931	A1641931 vq50e03.y
21	72	46.8	629	34	AA518112	AA518112 v123c06.r
22	70.6	45.8	386	37	AA692521	AA692521 vt59g08.r
23	69.8	45.3	452	39	AA822200	AA822200 vp36d05.r
24	68.6	44.5	554	36	AA623196	AA623196 v114b07.r
25	68.6	44.5	507	36	AA623217	AA623217 v114d07.r
26	68.6	44.5	207	39	AA869318	AA869318 vq50e03.r
27	65.4	42.5	310	31	AA297150	AA297150 EST112734
28	63	40.9	205	50	A1675991	A1675991 wd08g07.x
29	45	29.2	410	42	A111432	A111432 UT-R-YO-1
30	39.8	25.8	463	37	AA690350	AA690350 vt31a09.r
31	39.8	25.8	297	40	AA929159	AA929159 vq58g07.r
32	34.4	22.3	533	33	AA429919	AA429919 zw66f05.s
33	34.4	22.3	488	33	AA430055	AA430055 zw65h07.s
34	33.8	21.9	548	28	AA073158	AA073158 mm80f10.r
35	33.4	21.7	455	29	AA174903	AA174903 ms78e05.r
36	33.4	21.7	289	29	AA178274	AA178274 mt11g07.r
37	33.4	21.7	450	29	AA182290	AA182290 mt76a07.r
38	33.4	21.7	401	29	AA183456	AA183456 mt26f03.r
39	32.8	21.3	701	48	A1608718	A1608718 tv98a03.x
40	32.8	21.3	350	50	A1675394	A1675394 wb98h03.x
41	31.6	20.5	579	36	AA616701	AA616701 vn67e10.r
42	31.6	20.5	497	37	AA690345	AA690345 vt31a03.r
43	31.6	20.5	398	39	AA871707	AA871707 vq40a04.r
44	31.6	20.5	566	39	AA871728	AA871728 vq40c11.r
45	31.6	20.5	545	40	AA929451	AA929451 vt30g03.r

ALIGNMENTS

RESULT 1  
LOCUS AA296955 378 bp mRNA  
DEFINITION EST112726 Colon I Homo sapiens cDNA 5' end, mRNA sequence.  
ACCESSION AA296955  
NID g1949515  
VERSION AA296955.1 GI:1949515  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 378)  
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., White, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., Bilt, C.J., Sutton, G., Blake, J.A., Brandon, R.C., Man, M.-L., Clifton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, P., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geochagan, N.S., Glodex, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, F.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, R.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Fertie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Ye, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.  
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl.), 3-174 (1995)  
On Sep 12, 1996 this sequence version replaced gi:1288227.  
JOURNAL MEDLINE  
COMMENT Other ESTs: THC167738  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/cdb/ngi/ngi.html>)  
Seq primer: M3 Reverse.  
Location/Qualifiers  
1..378  
/organism="Homo sapiens"  
/db\_xref="ATCC (ihost):127834"  
/db\_xref="taxon:9606"  
/clone\_lib="Colon 1"  
/dev\_stage="adult"  
/note="Organ: colon; Vector: pBluescript SK-; Site: 1; Ecotri: site-2; XhoI"  
BASE COUNT 112 a 79 c 64 g 119 t 4 others  
ORIGIN

Query Match 100.0%; Score 154; DB 31; Length 378;  
Best Local Similarity 100.0%; Pred. No. 2.3e-37;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTCATTCGATTCAGCGCTGTGATTAAGTGCATTCGAAATCAGAAATATCC 60  
|||||  
DB 94 ACAGATCTTTTCATTCGATTCAGCGCTGTGATTAAGTGCATTCGAAATCAGAAATATCC 153  
|||||  
QY 61 AACATTGCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGACACCTAGTCTT 120  
|||||  
DB 154 AACATTGCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGACACCTAGTCTT 213  
|||||  
QY 121 GATGAAAGCTGCTGCTTGTCTTAATATTCATA 154  
|||||  
DB 214 GATGAAAGCTGCTGCTTGTCTTAATATTCATA 247  
|||||

RESULT 2  
AA508854 490 bp mRNA EST 18-AUG-1997  
LOCUS AA508854.1 NC1\_CGAP\_Co4 Homo sapiens cDNA clone IMAGE:968803 3',  
DEFINITION n123n10.s1

ACCESSION mRNA sequence.  
NID AA508854  
VERSION 92246357  
KEYWORDS GI:2246357  
SOURCE EST.  
ORGANISM human.  
REFERENCE Homo sapiens  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 490)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1397407.  
JOURNAL MEDLINE  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)  
Insert Length: 856 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 404.  
Location/Qualifiers  
1..490  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="968803"  
/clone\_lib="NC1\_CGAP\_Co4"  
/sex="pooled"  
/tissue\_type="colon"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker: 1st strand cDNA was prepared from pooled colon tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. This library is not normalized. Library constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 155 a 82 c 89 g 164 t  
ORIGIN

Query Match 100.0%; Score 154; DB 34; Length 490;  
Best Local Similarity 100.0%; Pred. No. 2.4e-37;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTCATTCGATTCAGCGCTGTGATTAAGTGCATTCGAAATCAGAAATATCC 60  
|||||  
DB 488 ACAGATCTTTTCATTCGATTCAGCGCTGTGATTAAGTGCATTCGAAATCAGAAATATCC 429  
|||||  
QY 61 AACATTGCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGACACCTAGTCTT 120  
|||||  
DB 428 AACATTGCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGACACCTAGTCTT 369  
|||||  
QY 121 GATGAAAGCTGCTGCTTGTCTTAATATTCATA 154  
|||||  
DB 368 GATGAAAGCTGCTGCTTGTCTTAATATTCATA 335  
|||||

RESULT 3  
A1660234 716 bp mRNA EST 10-MAY-1999  
LOCUS A1660234.1 Soares\_Dieckgraefe\_colon\_NHCD Homo sapiens cDNA clone  
DEFINITION web6902.x1

IMAGE:2346290 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA  
sequence.  
ACCESSION A1660234  
NID 94763804  
VERSION A1660234.1 GI:4763804  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 716)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949219.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40up from G1bco  
High quality sequence stop: 387.  
Location/Qualifiers  
1..716  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2346290"  
/clone.lib="Soares-Dieckgraefe.colon.NHCD"  
/tissue.type="colonic mucosa from 3 patients with Crohn's  
disease."  
/lab.host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pRT30-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACATCTGAGTGGAGGCGGCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pRT30 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieck@wustl.edu); colonic mucosa represents a range of  
disease involvement from moderate to severe Crohn's  
disease; samples include both perforating (fistulas) and  
non-perforating samples. Library constructed by Bento  
Soares and M. Fatima Bonaldo.

BASE COUNT 220 a 123 c 140 g 233 t  
ORIGIN

Query Match 100.0%; Score 154; DB 49; Length 716;  
Best Local Similarity 100.0%; Pred. No. 2.5e-37;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTCTATTCGCTATTCAGGCTGTGATGAAGTCGATCGATGGAATAGAAATATCC 60  
|||||  
DB 485 ACAGATCTTTTCTATTCGCTATTCAGGCTGTGATGAAGTCGATCGATGGAATAGAAATATCC 426  
|||||

QY 61 AACATGCAAGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCTT 120  
|||||  
DB 425 AACATGCAAGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCTT 366  
|||||

QY 121 GATGAACGCTGCTCTCTGCTGCTTAATATTATTA 154  
|||||  
DB 365 GATGAACGCTGCTCTCTGCTGCTTAATATTATTA 332  
|||||

RESULT 4  
A1660957 734 bp mRNA EST 10-MAY-1999  
LOCUS A1660957/c  
DEFINITION wF20d08.x1 Soares-Dieckgraefe.colon.NHUC Homo sapiens cDNA clone  
IMAGE:2351151 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA

sequence.  
ACCESSION A1660957  
NID 94764540  
VERSION A1660957.1 GI:4764540  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 734)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980033.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40up from G1bco  
High quality sequence stop: 477.  
Location/Qualifiers  
1..734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="X"  
/clone="IMAGE:2351151"  
/clone.lib="Soares-Dieckgraefe.colon.NHUC"  
/tissue.type="colonic mucosa from 5 ulcerative colitis  
patients."  
/lab.host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pRT30-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACATCTGAGTGGAGGCGGCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pRT30 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieck@wustl.edu); colonic mucosa represents a range of  
disease involvement from mild cryptitis to severe  
ulceration, fibrosis, and degeneration. Library  
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 200 a 143 c 159 g 228 t  
ORIGIN

Query Match 100.0%; Score 154; DB 49; Length 734;  
Best Local Similarity 100.0%; Pred. No. 2.5e-37;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTCTATTCGCTATTCAGGCTGTGATGAAGTCGATCGATGGAATAGAAATATCC 60  
|||||  
DB 281 ACAGATCTTTTCTATTCGCTATTCAGGCTGTGATGAAGTCGATCGATGGAATAGAAATATCC 222  
|||||

QY 61 AACATGCAAGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCTT 120  
|||||  
DB 221 AACATGCAAGATATCTTTGTTATTCCTCCACAGACTCCGCCAGAGACACTAGTCTT 162  
|||||

QY 121 GATGAACGCTGCTCTCTGCTGCTTAATATTATTA 154  
|||||  
DB 161 GATGAACGCTGCTCTCTGCTGCTTAATATTATTA 128  
|||||

RESULT 5  
A1721121 335 bp mRNA EST 10-JUN-1999  
LOCUS A1721121/c  
DEFINITION as73d08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
IMAGE:2334351 3' similar to TR:088826 088826 GDB-5 PROTEIN. ;, mRNA  
sequence.





ACCESSION D25727  
NTD 9500422  
VERSION D25727.1 GI:500422  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 255)  
AUTHORS Okubo,K., Yoshii,J., Yokouchi,H., Kameyama,M. and Matsubara,K.  
TITLE Global analysis of gene expression in colon mucosa: a large scale  
random cDNA sequencing analysis  
JOURNAL Unpublished (1994)  
COMMENT Contact: Okubo,K., Itoh,K., Yoshii,J., Yokouchi,H. and Matsubara,K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.  
FEATURES  
SOURCE Location/Qualifiers  
1..255  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cm2037"  
/clone.lib="Human colon mucosa"  
/note="Adult male, tissue\_type = colon mucosa"  
BASE COUNT 76 a 52 c 39 g 80 t 8 others  
ORIGIN

Query Match 72.7% Score 112: DB 20: Length 255;  
Best Local Similarity 97.4% Pred. No. 1.4e-24;  
Matches 112: Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 40 GATCTGAATCAGAAATATCCACATTCGACGAGTATCTTTTATTCCTCCACAGACT 99  
|||||  
Db 1 GATCTGAATCAGAAATATCCACATTCGACGAGTATCTTTTATTCCTCCACAGACT 60  
|||||

QY 100 CGGCCAGACACACTGCTCTGATGAAGCTGCTCTTCTTCTTAATTCATA 154  
|||||  
Db 61 CGGCCAGACACACTGCTCTGATGAAGCTGCTCTTCTTCTTAATTCATA 115  
|||||

RESULT 8  
LOCUS A1687981 392 bp mRNA EST 27-MAY-1999  
DEFINITION wa77f06.x1 Soares.NFL.T\_GBC.S1 Homo sapiens cDNA clone  
IMAGE:2302211 3' similar to TR:088826 O88826 GOB-5 PROTEIN. ;, mRNA  
sequence.  
ACCESSION A1687981  
NTD 94899275  
VERSION A1687981.1 GI:4899275  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 392)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2947973.  
CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco.  
FEATURES  
SOURCE Location/Qualifiers  
1..392  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="IMAGE:2302211"  
/clone.lib="Soares.NFL.T\_GBC.S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site: 1: Not I; Site 2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCL-CGAP-GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731359. Subtraction by Bento  
Soares and M. Fatima Bonaldo.  
BASE COUNT 125 a 69 c 67 g 131 t  
ORIGIN

Query Match 59.0% Score 90.8: DB 50: Length 392;  
Best Local Similarity 97.9% Pred. No. 4.4e-18;  
Matches 92: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 61 AACATGACGAGTACTTGTATTCCTCCACAGACTCCGACGACACTAGTCT 120  
|||||  
Db 392 AACATGACGAGTACTTGTATTCCTCCACAGACTCCGACGACACTAGTCT 333  
|||||

QY 121 GATGAACGCTGCTCTCTTCTTCTTAATTCATA 154  
|||||  
Db 332 GATGAACGCTGCTCTCTTCTTCTTAATTCATA 299  
|||||

RESULT 9  
LOCUS A1506262/c 175 bp mRNA EST 11-MAR-1999  
DEFINITION vq50e03.x1 Barsdale bowel MLR9 Mus musculus cDNA clone  
IMAGE:1097692 3' similar to SW:EGC\_BOVIN P54281 EPITHELIAL  
CHLORIDE CHANNEL PROTEIN ;, mRNA sequence.  
ACCESSION A1506262  
NTD 94404113  
VERSION A1506262.1 GI:4404113  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 175)  
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theisling,B., Allen,M., Bowers,Y.,  
Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R.,  
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
TITLE The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished (1999)  
COMMENT On May 18, 1998 this sequence version replaced gi:3136726.  
CONTACT: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG1:603924  
This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
Possible reversed clone: similarity on wrong strand  
High quality sequence stop: 1.  
FEATURES  
SOURCE Location/Qualifiers  
1..175  
/organism="Mus musculus"  
/strain="FVB/N"



```
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="105098"
/clone_lib="Barstead mouse irradiated colon MRLB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
15'TGTACGATCTGAGTGGAGGCGGCCCTTTTCTTTTCTTTTCTTTTCTTTT
T 3'1; double-stranded cDNA was ligated to Eco RI
adaptors (AATTCGATCTTG), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library constructed by Bob Barstead."
```

BASE COUNT  
ORIGIN

```
Query Match 47.8%; Score 73.6; DB 35; Length 418;
Best Local Similarity 81.7%; Pred. No. 7.8e-13;
Matches 85; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
OY 1 ACAGATCTTTCATTCCTATTCAGCTGTGATAGTGCATGAAATCAGAAATCC 60
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 66 ACAGATATCTCTATTCCTATTCAGCTGTGATAGTGCATGAAATCAGAAATCC 125
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 61 AACATGCAGAGATCTTGTATTTCCTCCACAGACTCCGCC 104
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 126 AACATGCAGAGGCTGTCTGTCTTCATCCCGCTCAGAGAGCCGCC 169
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 12
LOCUS AA688953 552 bp mRNA EST 12-DEC-1997
DEFINITION vs04h07.r1 Barstead mouse irradiated colon MRLB7 Mus musculus CDNA
clone IMAGE:113727 5' similar to SW:ECIC_BOVIN P54281 EPIITHELIAL
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.
```

```
ACCESSION AA688953
NID 92678382
VERSION AA688953.1 GI:2678382
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
```

```
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 552)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Nov 6, 1997 this sequence version replaced gi:930439.
```

CONTACT: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:616549  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 509.  
Location/Qualifiers  
1..552  
/organism="Mus musculus"

```
/strain="FVB/N"
/db_xref="taxon:10090"
/map="13"
/clone_image="113727"
/clone_lib="Barstead mouse irradiated colon MRLB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
15'TGTACGATCTGAGTGGAGGCGGCCCTTTTCTTTTCTTTTCTTTTCTTTT
T 3'1; double-stranded cDNA was ligated to Eco RI
adaptors (AATTCGATCTTG), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library constructed by Bob Barstead."
```

BASE COUNT  
ORIGIN

```
Query Match 47.8%; Score 73.6; DB 37; Length 552;
Best Local Similarity 81.7%; Pred. No. 8e-13;
Matches 85; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
OY 1 ACAGATCTTTCATTCCTATTCAGCTGTGATAGTGCATGAAATCAGAAATCC 60
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 253 ACAGATATCTCTATTCCTATTCAGCTGTGATAGTGCATGAAATCAGAAATCC 312
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 61 AACATGCAGAGATCTTGTATTTCCTCCACAGACTCCGCC 104
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 313 AACATGCAGAGGCTGTCTGTCTTCATCCCGCTCAGAGAGCCGCC 356
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 13
LOCUS AA689856 301 bp mRNA EST 16-DEC-1997
DEFINITION vt62c11.r1 Barstead mouse irradiated colon MRLB7 Mus musculus CDNA
clone IMAGE:116768 5' similar to SW:ECIC_BOVIN P54281 EPIITHELIAL
CHLORIDE CHANNEL PROTEIN ; mRNA sequence.
```

```
ACCESSION AA689856
NID 92690792
VERSION AA689856.1 GI:2690792
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
```

```
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 301)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Apr 14, 1993 this sequence version replaced gi:693293.
```

CONTACT: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:635580  
Trace considered overall poor quality  
Possible reversed clone; similarity on wrong strand  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1..301  
/organism="Mus musculus"

## FEATURES

Location/Qualifiers

source

```
1. .421
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1138042"
/dev_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: p7773D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dT) primer
15'GTGTACGATCTGAAGTGGAGCGGCGCCCTTTTCTTTTCTTTTCTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors [ATTGGAATCCCTG], digested with Not I and cloned
into the Not I and Eco RI sites of the modified p7773
vector. Library constructed by Bob Barstead."
```

BASE COUNT 119 a 98 c 89 g 115 t

ORIGIN

```
Query Match 47.8%; Score 73.6; DB 37; Length 421;
Best Local Similarity 81.7%; Pred. No. 7.8e-13;
Matches 85; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 ACAGATCTTTTCATTGCTATTCAGGCTGTGATAGGTCATCTGAATCAGAAATATCC 60
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 29 ACAGATATCTTCATTGCTATTCAGGCTGTGATAGGTCATCTGAATCAGAAATATCC 88

QY 61 AACATTGCACGAGTATCTTGTATTATTCCTCCACAGAGCTCCGCC 104
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 89 AACATTGCACGAGTATCTTGTATTATTCCTCCACAGAGCTCCGCC 132
```

Search completed: August 6, 1999, 13:33:26  
Job time: 13004 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:04:15 ; Search time 921.96 Seconds

(without alignments)  
41.791 Million cell updates/sec

Title: US-09-049-696-15

Perfect score: 154  
Sequence: 1 ACAGATCTTTTCATATGCTAT.....TCCTGTCTCTAATATTCATA 154

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	92.9	878	1	T45884 Human colon specif
2	143	92.9	878	1	PV16672 Polynucleotide seq
3	112	72.7	255	1	T22483 Human gene signatu
4	29.4	19.1	3984	1	T42118 H-lats gene encodi
5	29.4	19.1	3213	1	T42119 M-lats gene encodi
6	29.4	19.1	1374	1	X33982 Human warts gene f
7	29.2	19.1	2442	1	X33981 Human warts protei
8	28.2	18.6	6327	1	T32301 Dermatophytosis sp
9	28.6	18.6	110000	1	V21209_08 Continuation (9 of
10	28.2	18.3	3662	1	X13152 Enterococcus faeca
11	28	18.2	110000	1	X20248_05 Continuation (6 of
12	28	18.2	674	1	X26841 Microspore-specifi
13	27.6	17.9	2385	1	T33590 Signal transducer
14	27.6	17.9	5368	1	V25978 Human heterogeneu
15	27.6	17.9	35099	1	V27112 Adenovirus 17, Rec
16	27.6	17.9	1554	1	V27114 Adenovirus 17 pent
17	27.6	17.9	3120	1	V71029 Green fluorescent
18	27.6	17.9	3138	1	V71041 Stat5-green floure
19	27.6	17.8	16870	1	X13035 Enterococcus faeca
20	27.4	17.8	2844	1	N71250 Sequence of Vicia
21	27.2	17.7	731	1	T09091 Potato U2B gene P
22	27	17.5	110000	1	V21209_03 Continuation (4 of
23	26.8	17.4	333	1	T19132 Human gene signatu
24	26.8	17.4	2115	1	T96642 Human TULP1 cDNA
25	26.8	17.4	7650	1	V82019 Moraxella catarrha
26	26.8	17.4	7642	1	V82020 Moraxella catarrha
27	26.6	17.3	4488	1	O51426 Human FACC cDNA c1
28	26.6	17.3	2344	1	O51427 Human FACC cDNA c1
29	26.6	17.3	3150	1	O51428 Human FACC cDNA c1
30	26.6	17.3	4567	1	X33945 Fanccon1 anaemia co
31	26.4	17.1	8959	1	O55133 Stephylococcus aur
32	26.4	17.1	2170	1	V52445 Streptococcus pneu
33	26.4	17.1	9623	1	V74477 Streptococcus aur
34	26.4	17.1	1560	1	X27725 Canine interleukin
35	26.4	17.1	1215	1	X27725 Canine interleukin
36	26.2	17.0	315	1	T25014 Human gene signatu
37	26.2	17.0	110000	1	V21209_09 Continuation (10 o
38	26.2	17.0	1378	1	X13473 Enterococcus faeca
39	26.2	17.0	430	1	X51984 Human secreted pro
40	26	16.9	267	1	O77391 Human genome fragm
41	25.8	16.8	8789	1	O86851 Human mitotin gene
42	25.8	16.8	10136	1	T34578 Kinetochore protei
43	25.8	16.8	8789	1	V09076 Mitotin nucleic ac

## ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB ID	Description
44	25.6	16.6	280	1	N60397	Sequence of glucoa
45	25.6	16.6	2634	1	X27062	S. tuberosum isoam
ALIGNMENTS						
RESULT 1	T45884	143	92.9	878	1	T45884 Human colon specif
AC	T45884	143	92.9	878	1	PV16672 Polynucleotide seq
DT	13-MAR-1997	(first entry)				
DE	Human colon specific gene CSG5 cDNA partial clone.					
KW	Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;					
KW	gene therapy; ss.					
OS	Homo sapiens.					
FT	Key	Location/Qualifiers				
FT	cds	2..691				
FT		/*tag= a				
PN	W09639419-A1.					
PD	12-DEC-1996.					
PF	06-JUN-1995; U07289.					
PR	06-JUN-1995; MO-U07289.					
PA	(HUMA-) HUMAN GENOME SCI INC.					
PI	Rosen CA, Yu G;					
DR	WPI: 97-043054/04.					
DR	P-PSDB: W06548.					
PT	Human colon specific genes and their expression products - detection					
PT	of which, in non-colon tissue samples, can be used as indication of					
PS	colon cancer metastasis					
PS	Claim 1: Fig 5; 60pp; English.					
CC	13 cDNA clones (T45880-92), most of them partial clones, correspond					
CC	to human colon specific genes, designated CSG1, CSG2, etc., that					
CC	are primarily expressed in tissues derived from the colon. CSG7					
CC	and CSG10 show reduced expression in colon cancer cells as compared					
CC	to that in normal cells; the remaining genes are overexpressed in					
CC	colon cancer. The partial cDNA sequences can be used to isolate					
CC	full-length clones and genomic clones including the complete gene.					
CC	CSG nucleic acids can be used to produce CSG polypeptides (see also					
CC	W06545-53) in transformed host cells, as probes to detect disorders					
CC	of the colon, partic. colon cancer and colon cancer metastasis, and					
CC	in gene therapy.					
CC	Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;					
SO	Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;					
Query Match						
Best local similarity 92.9%; Score 143; DB 1; Length 878;						
Matches 154; Conservative 0; Mismatches 0; Indels 1; Gaps 1;						
QY	1 ACAGATCTTTTCATATGCTAT.....TCCTGTCTCTAATATTCATA 154					
DB	545 ACAGATCTTTTCATATGCTAT.....TCCTGTCTCTAATATTCATA 154					
QY	61 AACATTCGACGAGTATCTTGTATTCCTCCACAGACCTCCGACGACCTAGTCT 120					
DB	605 AACATTCGACGAGTATCTTGTATTCCTCCACAGACCTCCGACGACCTAGTCT 664					
QY	121 GATGAACGCTGCTCTTGT-CCTAATATTCATA 154					
DB	665 GATGAACGCTGCTCTTGT-CCTAATATTCATA 699					
RESULT 2						
ID	V16672	143	92.9	878	1	V16672 standard; cDNA; 878 BP.
AC	V16672	143	92.9	878	1	PV16672 Polynucleotide seq
DT	22-JUN-1998	(first entry)				
DE	Polynucleotide sequence of a colon-specific gene.					
KW	Colon-specific gene; probe; detection; expression; human;					
KW	diagnostic assay; colon cancer; antibody; screening; ss.					
OS	Homo sapiens.					
FT	Key	Location/Qualifiers				
FT	cds	2..685				

```

TT      /*tag= a /note= "no stop codon given"
FT
PN      US5733748-A.
PF      31-MAR-1998.
PE      06-JUN-1995; 469667.
PR      06-JUN-1995; US-469667.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Rosen C, Yu G;
PR      WPI: 98-229823/20.
DR      p-PSDB; M46879.
PT      Colon-specific nucleic acids - useful as probes for detecting colon
PT      cancer micrometastases
PS      Claim 15; Fig 5A-B; 51pp; English.
CC      V16668-81 represent polynucleotide sequences of partial or full-length
CC      cDNA clones of colon-specific genes. The polynucleotides can be used
CC      as probes to detect expression of the corresponding human genes. e.g. in
CC      diagnostic assays for detecting micrometastases of colon cancer.
CC      Recombinant cells containing the polynucleotides can be used to
CC      produce the polypeptides, in order that antibodies can be raised and
CC      used in further screening or diagnostics.
CC      Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;
SQ
Query Match 92.9%; Score 143; DB 1; Length 878;
Best Local Similarity 99.4%; Pred No. 3,8e-37;
Matches 154; Conservative 0; Mismatches 0; Indels 1; Gaps 1.
OY      1 ACAGATCTTTGATTCATTCAGGCTGTGATAGTCGATCGAATCAGAAATATCC 60
DB      545 ACAGATCTTTTATTCCTATTCAGGCTGTGATAGTCGATCGAATCAGAAATATCC 604
OY      61 AACATTGCACAGATCTTTGTTATTCCTCCACAGATCCGCCAAGACACCTAGTCT 120
DB      605 AACATTGCACAGATCTTTGTTATTCCTCCACAGATCCGCCAAGACACCTAGTCT 664
OY      121 GATGAACGTCCTCCTTGT-CCTAATATTCATA 154
DB      665 GATGAACGTCCTCCTTGTGCTTGTGCTAATATTCATA 699
RESULT 3
T22483
ID      T22483 standard: cDNA to mRNA; 255 BP.
AC      T22483:
DT      22-AUG-1996 (first entry)
DE      Human gene signature HUMGS04094.
KW      Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW      human; cloning; mapping; non-biased library; diagnosis; detection;
KW      cell typing; abnormal cell function; ss.
OS      Homo sapiens.
PN      WO9514772-A1.
PD      01-JUN-1995.
PE      11-NOV-1994; J01916.
PR      12-NOV-1993; JP-355504.
PA      (MATS/) MATSUBARA K.
PA      (OKUB/) OKUBO K.
PI      Matsubara K, Okubo K;
PI      WPI: 95-206931/27
PT      Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT      for diagnosis of abnormal cell function, by preparing cDNA that
PT      reflects relative abundance of corresp. mRNA in specific human
PT      tissues
PS      Claim 1: Page 1138-1139; 2245pp; Japanese.
CC      A single-stranded DNA (or its complementary strand or the corresp.
CC      double-stranded DNA) which comprises one of the 7837 "G3" sequences
CC      given in T19001-T26837 and which is able to hybridise to part of
CC      human genomic DNA, cDNA or mRNA is claimed. The G3 (gene signature)
CC      sequences were obtained from 3'-directed cDNA libraries prepared
CC      from various human tissues; synthesis of cDNA was initiated from the
CC      3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC      untranslated sequence is unique to a particular mRNA species, almost
CC      all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC      is constructed so as to reflect accurately the relative abundance of

```

```

CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SO Sequence 255 BP; 76 A; 52 C; 39 G; 80 T;

Query Match 72.7%; Score 112; DB 1; Length 255;
Best Local Similarity 97.4%; Pred. No. 2,5e-27;
Matches 112; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 40 GATCTGAAATACGAAATATCCACATTCGACGAGTATCTTGTATTATCTCCACAGACT 99
DB 1 GATCTGAAATACGAAATATCCACATTCGACGAGTATCTTNTTATNCCTCCACAGACT 60
OY 100 CCGCAGAGACACCTAGTCTCGATGAAGAAGCTGCTCTTGTCTTAATATCAATA 154
DB 61 CCGCAGAGACACCTAGTCTCGATGAAGAAGCTGCTCTTGTCTTAATATCAATA 115

RESULT 4
T42118/c
ID T42118 standard; cDNA; 3984 BP.
AC T42118.
DF 31-JAN-1997 (first entry)
DE H-lats gene encoding large tumour suppressor.
KW Human; h-lats gene; large tumour suppressor; fetal brain;
KW protein-serine/threonine-kinase; cell proliferation; antisense;
KW dominant-negative; cancer; degenerative disorder; trauma;
KW growth deficiency; therapy; antitumour; vulnarney; diagnostic;
KW transgenic plant; transgenic animal; growth; senescence; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 231..3623
FT cds /tag= a
FT /product= lats protein
PD WO9630402-A1.
PD 03-OCT-1996.
PR 26-MAR-1996; U04101.
PR 27-MAR-1995; U5-41111.
PA (UYUA) UNIV YALE.
PI Tao W, Wang W, Xu T, Yu W, Zhang S;
DR WPI: 96-455275/45.
DR P-PSNB: W05178.
PT New isolated large tumour suppressor gene - used to develop prods.
PT for inhibiting cell proliferation or for enhancing proliferation
PS Claim 6; Page 118-123; 215pp; English.
CC This sequence encodes a human large tumour suppressor h-lats protein,
CC cDNA library using a 2.1-kb DNA probe from the mouse m-lats gene
CC (T42119). The full-length cDNA is present in plasmid pBS(KS)-h-lats.
CC The gene encodes a putative protein-serine/threonine-kinase, and
CC inhibits cell proliferation and plays a crucial role throughout
CC development. Activators or inhibitors of lats function (e.g. an
CC antisense oligonucleotide or dominant-negative lats fragment) may be
CC used in therapy of cancer or other proliferative disorders,
CC degenerative disorders, trauma, growth deficiency, etc., and
CC fragments of the gene may be used as diagnostic probes. A
CC lats-inhibitor sequence may be expressed in a transgenic plant or
CC farm animal to confer increased growth and inhibit senescence.
SQ Sequence 3984 BP; 1280 A; 847 C; 798 G; 1059 T;

Query Match 19.1%; Score 29.4; DB 1; Length 3984;
Best Local Similarity 66.7%; Pred. No. 2.7;
Matches 42; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 62 ACATTGACAGCATCTTTGTTATTCCTCCACAGACTCCGACAGACACCTAGTCTG 121
DB 2421 ACATTAAGCTTAGTATCTACTTTCTTGCTAGACAGACTTACCACAAATGCTCATATCCTA 2362
OY 122 ATG 124

```



DB 2361 GTC 2359

RESULT 5

T42119/c  
ID T42119 standard; CDNA: 3213 BP.  
AC T42119;  
DT 31-JAN-1997 (first entry)  
DE M-lats gene encoding large tumour suppressor.  
KW Mouse; m-lats gene; large tumour suppressor; fetal brain;  
KW Protein-serine/threonine-kinase; cell proliferation; antisense;  
KW dominant-negative; cancer; degenerative disorder; trauma;  
KW growth deficiency; therapy; antitumour; vulnerray; diagnostic;  
KW transgenic plant; transgenic animal; growth; senescence; ds.  
OS Mus musculus.  
FH Key  
FT Location/Qualifiers  
FT 1..2889  
FT /product= m-lats protein  
FT /tag= a  
FT /product= m-lats protein  
PN WO9630402-A1.  
PD 03-OCT-1996.  
PE 26-MAR-1996; U04101.  
PR 27-MAR-1995; US-41111.  
PA (UYVA) UNIV YALE.  
PI Tao W, Wang W, Xu T, Yu W, Zhang S;  
PI WPI: 96-453275/45.  
DR P-PSDB: W05179.  
PT New isolated large tumour suppressor gene - used to develop prods.  
PT for inhibiting cell proliferation or for enhancing proliferation  
PS Disclosure: Page 126-130: 215pp: English.  
CC This sequence encodes a mouse large tumour suppressor m-lats protein,  
CC and has been isolated from a newborn mouse brain phage lambda-ZAP  
CC cDNA library using a 2.2-kb DNA probe from the Drosophila lats gene  
CC (T42117). A homologous mouse sequence has also been isolated  
CC (m-lats2, T42120). The gene encodes a putative protein.  
CC serine/threonine-kinase, and inhibits cell proliferation and plays a  
CC crucial role throughout development. Activators or inhibitors of  
CC lats function (e.g. an antisense oligonucleotide or dominant-negative  
CC lats fragment) may be used in therapy of cancer or other  
CC proliferative disorders, degenerative disorders, trauma, growth  
CC deficiency, etc., and fragments of the gene may be used as diagnostic  
CC probes. A lats-inhibitor sequence may be expressed in a transgenic  
CC plant or farm animal to confer increased growth and inhibit  
CC senescence.  
SQ Sequence 3213 BP: 946 A; 762 C; 697 G; 808 T;

Query Match 19.1%; Score 29.4; DB 1; Length 3213;  
Best Local Similarity 66.7%; Pred. No. 2.5; 21; Indels 0; Gaps 0;  
Matches 42; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 62 ACATTGCAGAGTATCTTTATTCCTCCACAGACTCCGCCAGAGACCTAGTCTG 121  
DB 1687 ACAAGCCTTAGTATGACTTTCTTCTGCTAGACAGACTCCACCAACGCTCTATTCCTA 1628  
OY 122 ATG 124  
DB 1627 ATG 1625

RESULT 6  
X32982/c  
ID X32982 standard; DNA: 1374 BP.  
AC X32982;  
DT 30-JUN-1999 (first entry)  
DE Human warts gene fragment encoding a partial peptide fragment.  
KW Human; warts protein; Drosophila; protein kinase; epithelial cell;  
KW cell regulation; cell proliferation; cell differentiation; tumour; ss.  
OS Homo sapiens.  
PN WO9915558-A1.  
PD 01-APR-1999.  
PF 24-AUG-1998; J03739.

PR 24-SEP-1997; JP-258689.  
PA (SUME) SUMITOMO ELECTRIC IND CO.  
PI Kishimoto T, Nagamine Y, Nishiyama Y, Niwa S, Saya H;  
DR WPI: 99-244389/20.  
DR P-PSDB: Y06891.  
PT Gene of vertebrates homologous to a Drosophila gene  
PS Disclosure: Page 52-56; 68pp; Japanese.  
CC The invention relates to a human warts protein, homologous to Drosophila  
CC warts gene expression product. The warts expression product is a protein  
CC kinase involved in the regulation of proliferation and differentiation of  
CC epithelial cells and suppression of tumour formation. The gene, its  
CC expression product and antibodies are useful in the study of the  
CC mechanisms of tumour development. The present sequence represents a  
CC human warts gene fragment encoding a partial peptide fragment.  
SQ Sequence 1374 BP: 442 A; 252 C; 288 G; 392 T;

Query Match 19.1%; Score 29.4; DB 1; Length 1374;  
Best Local Similarity 66.7%; Pred. No. 1.9; 21; Indels 0; Gaps 0;  
Matches 42; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 62 ACATTGCAGAGTATCTTTATTCCTCCACAGACTCCGCCAGAGACCTAGTCTG 121  
DB 292 ACAAGCCTTAGTATGACTTTCTTCTGCTAGACAGACTCCACCAATGCTCTATTCCTA 233  
OY 122 ATG 124  
DB 232 GTC 230

RESULT 7  
X32981/c  
ID X32981 standard; DNA: 2442 BP.  
AC X32981;  
DT 30-JUN-1999 (first entry)  
DE Human warts protein encoding DNA.  
KW Human; warts protein; Drosophila; protein kinase; epithelial cell;  
KW cell regulation; cell proliferation; cell differentiation; tumour; ss.  
OS Homo sapiens.  
PN WO9915558-A1.  
PD 01-APR-1999.  
PF 24-AUG-1998; J03739.  
PR 24-SEP-1997; JP-258689.  
PA (SUME) SUMITOMO ELECTRIC IND CO.  
PI Kishimoto T, Nagamine Y, Nishiyama Y, Niwa S, Saya H;  
DR WPI: 99-244389/20.  
DR P-PSDB: Y06890.  
PT Gene of vertebrates homologous to a Drosophila gene  
PS Claim 7; Page 44-52; 68pp; Japanese.  
CC The invention relates to a human warts protein, homologous to Drosophila  
CC warts gene expression product. The warts expression product is a protein  
CC kinase involved in the regulation of proliferation and differentiation of  
CC epithelial cells and suppression of tumour formation. The gene, its  
CC expression product and antibodies are useful in the study of the  
CC mechanisms of tumour development. The present sequence represents a DNA  
CC encoding a human warts protein.  
SQ Sequence 2442 BP: 777 A; 528 C; 491 G; 646 T;

Query Match 19.1%; Score 29.4; DB 1; Length 2442;  
Best Local Similarity 66.7%; Pred. No. 2.3; 21; Indels 0; Gaps 0;  
Matches 42; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 62 ACATTGCAGAGTATCTTTATTCCTCCACAGACTCCGCCAGAGACCTAGTCTG 121  
DB 1360 ACAAGCCTTAGTATGACTTTCTTCTGCTAGACAGACTCCACCAATGCTCTATTCCTA 1301  
OY 122 ATG 124  
DB 1300 GTC 1298

RESULT 8



WP	Fragment Name	Begin	End	Accession
WP	X20248_00	1	110000	
WP	X20248_01	100001	210000	
WP	X20248_02	200001	310000	
WP	X20248_03	300001	410000	
WP	X20248_04	400001	510000	
WP	X20248_05	500001	610000	
WP	X20248_06	600001	710000	
WP	X20248_07	700001	810000	
WP	X20248_08	800001	910000	
WP	X20248_09	900001	910715	

FT CDS  
 FT 799..4307  
 FT /\*tag- f  
 FT /codon\_start=1  
 FT /product= "hnRNP A1 protein"  
 FT /note= "NCBI gi: 296650; contains introns"  
 FT 814..1376  
 FT /\*tag- g  
 FT /number= 1  
 FT 1377..1493  
 FT /\*tag- h  
 FT /number= 2  
 FT 1494..1788  
 FT /\*tag- i  
 FT /number= 2  
 FT 1789..1935  
 FT /\*tag- j  
 FT /number= 3  
 FT 1936..2083  
 FT /\*tag- k  
 FT /number= 3  
 FT 2084..2294  
 FT /\*tag- l  
 FT /number= 4  
 FT 2295..2387  
 FT /\*tag- m  
 FT /number= 4  
 FT 2388..2480  
 FT /\*tag- n  
 FT /number= 5  
 FT 2481..2566  
 FT /\*tag- o  
 FT /number= 5  
 FT 2567..2659  
 FT /\*tag- p  
 FT /number= 6  
 FT 2660..2793  
 FT /\*tag- q  
 FT /number= 6  
 FT 2794..2868  
 FT /\*tag- r  
 FT /number= 7  
 FT 2869..3805  
 FT /\*tag- s  
 FT /number= 7  
 FT 3806..3961  
 FT /\*tag- t  
 FT /number= 8  
 FT 3962..4251  
 FT /\*tag- u  
 FT /number= 8  
 FT 4252..4311  
 FT /\*tag- v  
 FT /number= 9  
 FT 4312..4542  
 FT /\*tag- w  
 FT /number= 9  
 FT 4543..5240  
 FT /\*tag- x  
 FT /number= 10  
 FT  
 PN J10023893-A.  
 PD 27-JAN-1998.  
 PR 09-JUL-1996; 179521.  
 PR 09-JUL-1996; JP-179521.  
 PA (HITA) HITACHI LTD.  
 WP1: 98-152795/14.  
 DR PPSDB: W55828.  
 DR Detection of protein code region on DNA base sequence - using method  
 PT which reduces interference from noncoding region  
 PS Disclosure; Fig 2-5; 16pp; Japanese.  
 CC The present sequence encodes the human heterogeneous nuclear  
 CC ribonucleoprotein (hnRNP) core protein A1, which is used to exemplify  
 CC the method of the present invention. The hnRNP A1 gene can be found on  
 CC the genbank database accession number X12671, NID 932344. The present

CC invention describes a method for the detection of a protein coding  
 CC region in a DNA base sequence by judging the probability where a  
 CC specific sequence will be found. The method comprises utilizing the  
 CC appearance frequency in the coding and the noncoding regions. The  
 CC method allows noise in the noncoding region to be reduced.  
 SQ Sequence 5368 BP; 1476 A; 1052 C; 1270 G; 1570 T;  
 Query Match 17.9%; Score 27.6; DB 1; Length 5368;  
 Best Local Similarity 53.8%; Pred. No. 12;  
 Matches 57; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 OY 14 TTGCTATTCAGCGTGTGATAGTCGATCGAATCAGAAATATTCACATTCGACAG 73  
 DB 2713 TTTCATGCAAGATGATGAGAGTCAAGACTAGTGTAGATTAACCTATCATCCACGAC 2654  
 OY 74 TATCTTGTATTTCCTCCACAGACTCCGCCAGACAGACCTAGTCC 119  
 DB 2653 CACTGAAAGTTCTCTCACGACGAGAGTGTATCCACCGAAACC 2608  
 RESULT 15  
 ID V27112/c  
 V27112 standard; DNA; 35099 BP.  
 AC V27112;  
 DT 28-SEP-1998 (first entry)  
 DE Adenovirus 17.  
 KW Adeno virus 17; Ad17; vector; gene therapy; cystic fibrosis;  
 KM alpha-antitrypsin deficiency; respiratory disease; ss.  
 OS Mastadenovirus 17.  
 PN M09822609-A1.  
 PD 28-MAY-1998.  
 PE 20-NOV-1997; U21494.  
 PR 20-NOV-1996; US-752760.  
 PA Armentano DE, Gregory RJ, Smith AE;  
 PI (GEN2) GENZYME CORP.  
 DR WPI: 98-312493/27.  
 PT Recombinant adenovirus type 2 vector - useful for targeting  
 PT biologically active proteins, used to, e.g. treat cystic fibrosis  
 PS Disclosure, page 28-38; 67pp; English.  
 CC This is the complete nucleotide (nt) sequence of adenovirus 17  
 CC (Ad17). A claimed chimeric adenoviral vector comprises a nt  
 CC sequence of an adenovirus in which all or part of a gene encoding  
 CC a protein facilitating attachment to, or internalisation into, a  
 CC target mammalian cell is replaced by all or part of the  
 CC corresponding gene for a second adenovirus from subgroup D,  
 CC preferably selected from Ad9, Ad15, Ad17, Ad19, Ad20, Ad22, Ad26,  
 CC Ad27, Ad28, Ad30 and Ad39. The vector further comprises a  
 CC transgene operably linked to a eukaryotic promoter to allow  
 CC expression in a mammalian cell. The replaced gene preferably  
 CC encodes an Ad fibre such as Ad17 penton base (see V27113) and/or an Ad  
 CC penton base such as Ad17 penton base (see V27114). The chimeric  
 CC vector is used to target biologically active proteins to airway  
 CC epithelial cells, especially for gene therapy of diseases such as  
 CC cystic fibrosis or alpha-antitrypsin deficiency. The vectors are  
 CC based on the discovery that proteins from the subgroup D viruses,  
 CC a group not normally associated with human respiratory diseases,  
 CC can effectively bind and internalise within human airway  
 CC epithelial cells.  
 SQ Sequence 35099 BP; 7988 A; 9877 C; 9978 G; 7252 T;  
 Query Match 17.9%; Score 27.6; DB 1; Length 35099;  
 Best Local Similarity 60.8%; Pred. No. 22;  
 Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
 OY 68 CACGAGTATCTTGTATTATCTCCACAGACTCCGCCAGAGACCTAGTCCATGATAA 127  
 DB 14260 CAGCGGGCTCTGTGAAAGTCCACCCGAGCCGCGACAGACACACGCTCGGCTGAAA 14201  
 OY 128 CGTCTGCTCTCTGT 141  
 DB 14200 GGCTCTGTAGTGT 14187

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:22:56 ; Search time 3059.95 Seconds  
(without alignments)  
159.868 Million cell updates/sec

Title: US-09-049-696-14

Perfect score: 248  
Sequence: 1 ACCGAGAGCGGAAATTAC.....TTGAAAATGCACAGATCTT 248

Scoring table:

Searched: 2546578 seqs, 98626752 residues

Database :

- EST:\*
- 1: em\_est1.\*
  - 2: em\_est2.\*
  - 3: em\_est3.\*
  - 4: em\_est4.\*
  - 5: em\_est5.\*
  - 6: em\_est6.\*
  - 7: em\_est7.\*
  - 8: em\_est8.\*
  - 9: em\_est9.\*
  - 10: em\_est10.\*
  - 11: em\_est11.\*
  - 12: em\_est12.\*
  - 13: em\_est13.\*
  - 14: em\_est14.\*
  - 15: em\_est15.\*
  - 16: em\_est16.\*
  - 17: em\_est17.\*
  - 18: em\_est18.\*
  - 19: em\_est19.\*
  - 20: gb\_est1.\*
  - 21: gb\_est2.\*
  - 22: gb\_est3.\*
  - 23: gb\_est4.\*
  - 24: gb\_est5.\*
  - 25: gb\_est6.\*
  - 26: gb\_est7.\*
  - 27: gb\_est8.\*
  - 28: gb\_est9.\*
  - 29: gb\_est10.\*
  - 30: gb\_est11.\*
  - 31: gb\_est12.\*
  - 32: gb\_est13.\*
  - 33: gb\_est14.\*
  - 34: gb\_est15.\*
  - 35: gb\_est16.\*
  - 36: gb\_est17.\*
  - 37: gb\_est18.\*
  - 38: gb\_est19.\*
  - 39: gb\_est20.\*
  - 40: gb\_est21.\*
  - 41: gb\_est22.\*
  - 42: gb\_est23.\*
  - 43: gb\_est24.\*
  - 44: gb\_est25.\*
  - 45: gb\_est26.\*
  - 46: gb\_est27.\*
  - 47: gb\_est28.\*
  - 48: gb\_est29.\*
  - 49: gb\_est30.\*
  - 50: gb\_est31.\*
  - 51: gb\_est32.\*
  - 52: em\_est20.\*
  - 53: em\_est21.\*

- 54: em\_est22.\*
- 55: em\_est23.\*
- 56: em\_est24.\*
- 57: em\_est25.\*
- 58: em\_est26.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	248	100.0	734	49	A1660957 w120d08.x
C 2	165	66.5	501	51	A1721275 as82h08.x
C 3	137	55.2	310	31	AA297150 EST112734
C 4	84	33.9	378	31	AA296955 EST112726
C 5	58	23.4	335	51	A1721121 as73d08.x
C 6	55	22.2	716	49	A1660234 w6B8d02.x
C 7	19	7.7	248	20	T12464 B029F Heart
C 8	19	7.7	489	22	R72678
C 9	19	7.7	295	25	D79803
C 10	19	7.7	479	26	W65383
C 11	19	7.7	438	28	AA058647 zK69g06.r
C 12	19	7.7	487	29	AA148794 z104g01.r
C 13	19	7.7	349	32	AA376662 EST89286
C 14	19	7.7	584	43	AA159509 v2R0a07.r
C 15	19	7.7	589	47	AA042544 AA042544
C 16	19	7.7	438	47	A1498319
C 17	19	7.7	380	50	A1673384 w118e01.x
C 18	19	7.7	699	53	AA751772 96AS0605
C 19	18	7.3	220	38	AA751772 96AS0605
C 20	18	7.3	430	40	AA974527 cF13d09.s
C 21	17	6.9	301	21	F00550
C 22	17	6.9	398	22	H06548
C 23	17	6.9	388	22	H06578
C 24	17	6.9	433	22	R23291 yH28a06.r1
C 25	17	6.9	460	22	R56596 yG91h09.r1
C 26	17	6.9	430	22	R82767 y125f01.r1
C 27	17	6.9	353	23	H21570 y133a03.r1
C 28	17	6.9	348	25	N62174 yz62h03.s1
C 29	17	6.9	392	26	W18488 mb57b10.r1
C 30	17	6.9	1286	26	W42216
C 31	17	6.9	341	26	W56842 zc97d05.r1
C 32	17	6.9	409	26	W92717
C 33	17	6.9	486	27	AA010171 z617g11.r
C 34	17	6.9	360	28	AA083989 zn16h08.r
C 35	17	6.9	196	30	AA262131 z623b09.r
C 36	17	6.9	455	31	AA287914 z655b06.s
C 37	17	6.9	403	31	AA309604 EST180519
C 38	17	6.9	404	31	AA311238 EST180213
C 39	17	6.9	456	31	AA313265 EST185184
C 40	17	6.9	306	31	AA318698 EST20856
C 41	17	6.9	294	32	AA339049 EST44104
C 42	17	6.9	290	32	AA346229 EST52340
C 43	17	6.9	296	32	AA368657 EST79943
C 44	17	6.9	602	36	AA603946 DP27h03.s
C 45	17	6.9	503	36	AA616012 v095g09.r

ALIGNMENTS

RESULT 1  
A1660957/c A1660957 734 bp mRNA EST 10-MAY-1999  
LOCUS w120d08.x1 Soares,Dieckgraefe-colon\_NHUC Homo sapiens CDNA clone  
DEFINITION IMAGE:2351151.3/ similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA  
ACCESSION A1660957  
NID 94764540

VERSION A1660957.1 GI:4764540  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 734)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980033.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -400P from GIBCO  
High quality sequence stop: 477.  
Location/Qualifiers  
1..734  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="x"  
/clone="IMAGE:2351151"  
/clone\_lib="Soares.Dieckgraefe.colon\_NHUC"  
/tissue\_type="colonic mucosa from 5 ulcerative colitis patients"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: 1; Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGGAGCGCGCGAGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from mild cryptitis to severe ulceration, fibrosis, and degeneration. Library constructed by Bento Soares and M. Fatima Bonaldo.  
BASE COUNT 200 a 143 c 159 g 228 t 4 others  
ORIGIN  
Query Match 100.0%; Score 248; DB 49; Length 734;  
Best Local Similarity 100.0%; Pred. No. 3.7e-129; Indels 0; Gaps 0;  
Matches 248; Conservative 0; Mismatches 0;  
1 ACCTGAAGCGGAATTCACGGGGGCACTCATTAATCTGACAGCTCCTGGGG 60  
|||||  
Db 520 ACCGTAAAGCGGAATTCACGGGGGCACTCATTAATCTGACAGCTCCTGGGG 461  
Oy 61 ATGATTATGACCATGAAACAGCTCAAAAGTATATCATTCGATAAGTCAAGTATCTTG 120  
|||||  
Db 460 ATGATTATGACCATGAAACAGCTCAAAAGTATATCATTCGATAAGTCAAGTATCTTG 401  
Oy 121 ATCTAGAGCAAGTTCATGTAATCTCTCAAGTAAATCACTGCTCATCCCAAGG 180  
|||||  
Db 400 ATCTAGAGCAAGTTCATGTAATCTCTCAAGTAAATCACTGCTCATCCCAAGG 341  
Oy 181 AAGCCAACCTGAGGAGAGCTTTTGTGTTAAACGAGAAAACATTAAGTGGCA 240  
|||||  
Db 340 AAGCCAACCTGAGGAGAGCTTTTGTGTTAAACGAGAAAACATTAAGTGGCA 281  
Oy 241 CAGATCTT 248  
|||||  
Db 280 CAGATCTT 273  
RESULT 2

A1721275/c 501 bp mRNA EST 10-JUN-1999  
LOCUS A1721275  
DEFINITION as82h08.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
IMAGE:2335263 3' similar to SW:ECIC\_BOVIN P54281 EPIHEDIAL  
CHLORIDE CHANNEL PROTEIN;; mRNA sequence.  
ACCESSION A1721275  
NID 95038531  
VERSION A1721275.1 GI:5038531  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 501)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,  
Krisman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F.,  
Theisling, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246918.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watsn.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -400P from GIBCO  
High quality sequence stop: 394.  
Location/Qualifiers  
1..501  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="Xp11.2"  
/clone="IMAGE:2335263"  
/clone\_lib="Barstead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: 1; EcoRI; Site 2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGGAGCGCGCGAGCTTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [5' AATTCAGTAAAT 3' and 5' ATTCAGAG 3'], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bob Barstead."  
BASE COUNT 158 a 83 c 106 g 154 t  
ORIGIN  
Query Match 66.5%; Score 165; DB 51; Length 501;  
Best Local Similarity 99.5%; Pred. No. 1.7e-82;  
Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 33 ATTAATCTGAGTGGAGAGCTCTGGGGATGATTATGACATGAAACAGCTCAAGTAT 92  
|||||  
Db 501 ATTAATCTGAGTGGAGAGCTCTGGGGATGATTATGACATGAAACAGCTCAAGTAT 442  
Oy 93 ATCAATGGAATAGTACAAATATCTTCTGATCTAGAGACAAAGTCAATGATCTTCAA 152  
|||||  
Db 441 ATCAATGGAATAGTACAAATATCTTCTGATCTAGAGACAAAGTCAATGATCTTCAA 382  
Oy 153 GTGAATACTAGTCTCATCCCAAGAGCAACACTCTGAGGAAGCTTTTGTAA 212  
|||||  
Db 381 GTGAATACTAGTCTCATCCCAAGAGCAACACTCTGAGGAAGCTTTTGTAA 322  
Oy 213 CCAGAAAACATTAAGTGGCAAGATCTT 248

Db 321 CCAGAAACATTTACTTTGAAATGACAGATCTT 286

RESULT 3  
LOCUS AA297150 310 bp mRNA EST 18-APR-1997  
DEFINITION EST112734 Colon 1 Homo sapiens cDNA 5' end, mRNA sequence.  
ACCESSION AA297150  
NID 91949524  
VERSION AA297150.1 GI:1949524  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 310)  
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulcher,R.A.,  
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,  
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,  
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,  
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,  
Small,R.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,G.A.,  
Dincke,D., Feng,D.-F., Fertile,A., Fischer,C., Hastings,G.A.,  
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meisner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M., and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
On Sep 12, 1996 this sequence version replaced gi:1288566.  
Other ESTs: THC167738  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/cdb/hgi/hgi.html>)  
Seq primer: M3 Reverse.  
Location/Qualifiers  
1..310  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):127942"  
/db\_xref="taxon:9606"  
/clone\_1lb="Colon 1"  
/dev\_stage="adult"  
/note="Organ: colon; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"  
BASE COUNT 97 a 61 c 58 g 91 t 3 others  
ORIGIN

Query Match 55.2%; Score 137; DB 31; Length 310;  
Best Local Similarity 99.2%; Pred. No. 9.8e-67;  
Matches 237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 10 CGGAATTCACGGGGGAGCTCATTAATCTGAGACAGCTCTGGGATGATTATG 69  
DB 1 CGGAATTCACGGGGGAGCTCATTAATCTGAGACAGCTCTGGGATGATTATN 60  
QY 70 ACCATGGAACAGCTCACAAGTATATTCATTCGATTAAGTACAAATATTCGATCTCAGAG 129

Db 61 ACCATGGAACAGCTCACAAGTATATTCATTCGATTAAGTACAAATATTCGATCTCAGAG 120  
QY 130 ACAAGTTCAATGATCTCTTCAAGTGAATCTACTGCTCTCAATCCCAAGAACCACT 189  
DB 121 ACAAGTTCAATGATCTCTTCAAGTGAATCTACTGCTCTCAATCCCAAGAACCACT 180  
QY 190 CTGAGGAAGCTCTTTTGTGTTAAACAGAAACATTTCTTTGAAAATGGCACAGATCTT 248  
DB 181 CTGAGGAAGCTCTTTTGTGTTAAACAGAAACATTTCTTTGAAAATGGCACAGATCTT 239

RESULT 4  
LOCUS AA296955 378 bp mRNA EST 18-APR-1997  
DEFINITION EST112726 Colon 1 Homo sapiens cDNA 5' end, mRNA sequence.  
ACCESSION AA296955  
NID 91949515  
VERSION AA296955.1 GI:1949515  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 378)  
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulcher,R.A.,  
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,  
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,  
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,  
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,  
Small,R.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,G.A.,  
Dincke,D., Feng,D.-F., Fertile,A., Fischer,C., Hastings,G.A.,  
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meisner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M., and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
On Sep 12, 1996 this sequence version replaced gi:1288227.  
Other ESTs: THC167738  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/cdb/hgi/hgi.html>)  
Seq primer: M3 Reverse.  
Location/Qualifiers  
1..378  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):127834"  
/db\_xref="taxon:9606"  
/clone\_1lb="Colon 1"  
/dev\_stage="adult"  
/note="Organ: colon; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"  
BASE COUNT 112 a 79 c 64 g 119 t 4 others  
ORIGIN

Query Match 33.9%; Score 84; DB 31; Length 378;  
QY 33.9%; Score 84; DB 31; Length 378;

Mon Aug 9 13:26:50 1999

us-09-049-696-14\_1.rst

Page 4

Best Local Similarity 100.0%; Pred. No. 5.8e-37; Indels 0; Gaps 0;  
Matches 84; Conservative 0; Mismatches 0;

165 GCTCTCATCCCAAGAGCAAGCACTGTGAGAGCTTTTGTGTTAAACGAGAAACATT 224  
LOCUS |||||||  
Db 19 GCTCTCATCCCAAGAGCAAGCAAGCACTGTGAGAGCTTTTGTGTTAAACGAGAAACATT 78  
OY 225 ACTTTGAAATGCGACAGATCTT 248  
Db 79 ACTTTGAAATGCGACAGATCTT 102

RESULT 5  
A1721121 335 bp mRNA EST 10-JUN-1999  
LOCUS as73d08.x1 Barstead colon HPLMB7 Homo sapiens CDNA clone  
DEFINITION IMAGE:2334351 3' similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA  
sequence.

ACCESSION A1721121  
NID 95038377  
VERSION A1721121  
KEYWORDS GI:5038377  
SOURCE EST.  
ORGANISM human.

REFERENCE Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 335)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Getse, G., Jost, S.,  
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
Unpublished (1997)  
WashU-NCI human EST Project

TITLE Unpublished (1997)  
JOURNAL On Jun 22, 1998 this sequence version replaced gi:3246762.  
COMMENT

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from G1BCO  
High quality sequence stop: 312.  
Location/Qualifiers

FEATURES  
source 1..335  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2334351"  
/clone\_lib="Barstead colon HPLMB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st  
strand CDNA was primed with a Not I - oligo(dT) primer (5'  
TGTTGAGATCTGAGTGGAGCGGCGCCCTTTTGTGTTAAACGAGAAACATT  
3'); double-stranded CDNA was ligated to Eco RI adaptors  
(5' ATTACGATGAT 3' and 5' ATTACGATG 3'), digested  
with Not I and cloned into the Not I and Eco RI sites of  
the modified pT73 vector. Library constructed by Bob  
Barstead."

BASE COUNT 98 a 57 c 71 g 109 t  
ORIGIN

Query Match 23.4%; Score 58; DB 51; Length 335;  
Best Local Similarity 100.0%; Pred. No. 2.4e-22;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 191 TGAGAGAGCTTTTGTGTTAAACGAGAAACATTACTTTGAAATGCGACAGATCTT 248

Db 324 TGAGAGAGCTTTTGTGTTAAACGAGAAACATTACTTTGAAATGCGACAGATCTT 267

RESULT 6  
A1660234 716 bp mRNA EST 10-MAY-1999  
LOCUS we6802.x1 Soares, Dieckgraefe, colon\_NHCD Homo sapiens CDNA clone  
DEFINITION IMAGE:2346290 3' similar to TR:088826 088826 GDB-5 PROTEIN. ; mRNA  
sequence.

ACCESSION A1660234  
NID 94763804  
VERSION A1660234.1  
KEYWORDS GI:4763804  
SOURCE EST.  
ORGANISM human.

REFERENCE Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 716)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL On Mar 10, 1998 this sequence version replaced gi:2949219.  
COMMENT

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from G1BCO  
High quality sequence stop: 387.  
Location/Qualifiers

FEATURES  
source 1..716  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2346290"  
/clone\_lib="Soares, Dieckgraefe, colon\_NHCD"  
/tissue\_type="colonic mucosa from 3 patients with Crohn's  
disease"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand CDNA was primed with a Not I - oligo(dT) primer (5'  
TGTTGAGATCTGAGTGGAGCGGCGCCCTTTTGTGTTAAACGAGAAACATT  
3'); double-stranded CDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
St. Louis, MO). Colonic mucosa represents a range of  
disease involvement from moderate to severe Crohn's  
disease; samples include both perforating (fistulas) and  
non-perforating samples. Library constructed by Benito  
Soares and M. Fatima Bonaldo."

BASE COUNT 220 a 123 c 140 g 233 t  
ORIGIN

Query Match 22.2%; Score 55; DB 49; Length 716;  
Best Local Similarity 100.0%; Pred. No. 1e-20; Indels 0; Gaps 0;  
Matches 55; Conservative 0; Mismatches 0;

OY 165 GCTCTCATCCCAAGAGCAAGCACTGTGAGAGCTTTTGTGTTAAACGAGAA 219  
Db 560 GCTCTCATCCCAAGAGCAAGCAAGCACTGTGAGAGCTTTTGTGTTAAACGAGAA 506

RESULT 7  
T12464/c T12464 248 bp mRNA EST 28-NOV-1994  
LOCUS B029F Heart Homo sapiens CDNA clone B029, mRNA sequence.  
DEFINITION T12464  
ACCESSION



NID 9597151  
 VERSION T12464.1 GI:597151  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 248)  
 AUTHORS Llew,C.C., Hwang,D.M., Fung,Y.W., Laurensen,C., Cukerman,E.,  
 Tsui,S. and Lee,C.Y.  
 TITLE A catalogue of genes in the cardiovascular system as identified by  
 expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91, 10645-10649 (1994)  
 MEDLINE 95024171  
 COMMENT On Sep 21, 1992 this sequence version replaced gi:276460.

CONTACT: Llew CC  
 Department of Laboratory Medicine and Pathobiology  
 University of Toronto  
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5  
 Tel: 416/9788758  
 Fax: 416/9785650  
 Email: llew@cutcc.utoronto.ca  
 Seq primer: GGTGGCAGACTCCTCGAGCC.

FEATURES  
 source  
 1..248  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="B029"  
 /clone\_lib="Heart"  
 /lab\_host="E.coli Y1090"  
 /note="Vector: lambda gtl1; site\_1: EcoRI; site\_2: EcoRI"

BASE COUNT 76 a 65 c 35 g 72 t

ORIGIN

Query Match 7.7%; Score 19; DB 20; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GATGATTATGACCATGGAA 78  
 ||||||||||||||||  
 DB 141 GATGATTATGACCATGGAA 123

RESULT 8  
 LOCUS R72678 489 bp mRNA EST 02-JUN-1995  
 DEFINITION Y193d01.r1 Soares breast 2nbhbst Homo sapiens CDNA clone  
 IMAGE:156289 5', mRNA sequence.  
 ACCESSION R72678  
 NID 9846710  
 VERSION R72678.1 GI:846710  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 489)  
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
 Trevisan,S., E., Waterston,K., Williamson,A., Wohlmann,P. and  
 Wilson,R.  
 TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT On May 9, 1995 this sequence version replaced gi:802418.

CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu  
 Insert Size: 1114  
 High quality sequence stops: 373  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 1114 Std Error: 0.00  
 Seq primer: M13RP1  
 High quality sequence stop: 373.

FEATURES  
 source  
 1..489  
 /organism="Homo sapiens"  
 /db\_xref="GDB:569859"  
 /db\_xref="taxon:9606"  
 /clone\_image:156289  
 /clone\_lib="Soares breast 2NBHST"  
 /sex="Female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: breast; Vector: pTZ19 (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer (5'  
 TGTATACCAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3').  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of a modified pTZ19 vector (Pharmacia).  
 Library went through one round of normalization to a Cot -  
 230. Library constructed by Bento Soares and M. Fatima  
 Bonafido."

BASE COUNT 131 a 80 c 109 g 168 t 1 others

ORIGIN

Query Match 7.7%; Score 19; DB 22; Length 489;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GATGATTATGACCATGGAA 78  
 ||||||||||||||||  
 DB 369 GATGATTATGACCATGGAA 387

RESULT 9  
 LOCUS D79803 295 bp mRNA EST 09-FEB-1996  
 DEFINITION HUM329B03B Human aorta polyA+ (TFujiiwara) Homo sapiens CDNA clone  
 GEN:329B03 5', mRNA sequence.  
 ACCESSION D79803  
 NID G1180154  
 VERSION D79803.1 GI:1180154  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 295)  
 AUTHORS Fujiiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,  
 Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,  
 Takachi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,  
 Maekawa,H., Shin,S. and Nakamura,Y.  
 TITLE Fujiiwara et al. (1995)  
 JOURNAL Unpublished (1995)  
 COMMENT On Sep 21, 1992 this sequence version replaced gi:275986.

CONTACT: Tsutomu Fujiiwara  
 Otsuka GEN Research Institute  
 Otsuka Pharmaceutical Co., Ltd  
 463-10 Kagasuno Kawanuchi-cho, Tokushima, Tokushima, 771-01 Japan  
 Tel: 0886-65-2888  
 Fax: 0886-37-1035.

FEATURES  
 source  
 1..295  
 /organism="Homo sapiens"





RESULT 14  
 A1159509 584 bp mRNA EST 02-OCT-1998  
 LOCUS  
 DEFINITION V280a07.r1 Soares mouse mammary gland NbMNG Mus musculus cDNA clone  
 IMAGE:1332756 5', mRNA sequence.  
 A1159509  
 ACCESSION  
 NID 93692691  
 VERSION A1159509.1 GI:3692691  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE The WashU-HMNI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2152227.  
 Contact: Marra M/Mouse EST Project  
 WashU-HMNI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LBNL; contact the  
 IMAGE Consortium (info@image.lbnl.gov) for further information.  
 MGI:692300  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 514.  
 Location/Qualifiers  
 1..584  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /map="22: 4p16.3: 1"  
 /clone="IMAGE:1332756"  
 /clone\_lib="Soares mouse mammary gland NbMNG"  
 /sex="male"  
 /tissue\_type="mammary gland"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Organ: mammary gland; Vector: pT73D-Pac  
 (Pharmacia) with a modified polylinker; Site 1: Not I;  
 Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer [5',  
 TGTACCAATCTGAGTGGAGCGCGCGGATGTTTGTGTGTGTGTGTGTGTGTGT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia) digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. library  
 constructed and normalized by Bento Soares and M.Felina  
 Bonaldo."
 BASE COUNT 180 a 91 c 134 g 179 t  
 ORIGIN  
 Query Match 7.7%; Score 19; DB 43; Length 584;  
 Best Local Similarity 100.0%; Pred.No.1.8; 0; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;  
 Oy 215 AGAAACATTTCTTTGAA 233  
 ||||||||||||||||||  
 Db 379 AGAAACATTTCTTTGAA 397

RESULT 15  
 A0042544

LOCUS  
 DEFINITION A0042544 589 bp mRNA EST 04-DEC-1998  
 A0042544 Mouse four-cell-embryo cDNA Mus musculus cDNA clone  
 J1022E07 3', mRNA sequence.  
 A0042544  
 ACCESSION  
 NID 93956779  
 VERSION A0042544.1 GI:3956779  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS Ko,M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,  
 Depalma,G.E., Liang,Y., Kargul,G.J., Sharara,R., Lim,M.K. and  
 Dol,H.  
 TITLE Systematic analyses of genes expressed in 4-cell mouse embryo (the  
 ERATO/Dol Project at Wayne State University)  
 JOURNAL Unpublished (1998)  
 COMMENT On Jan 17, 1998 this sequence version replaced gi:2043790.  
 Contact: Hirofumi Dol  
 Dol Biosymmetry Project, ERATO  
 Japan Science and Technology Corporation (JST)  
 WBC Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan  
 Email: hdebia.1st.go.jp.  
 Location/Qualifiers  
 1..589  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /map="11"  
 /clone="J1022E07"  
 /clone\_lib="Mouse four-cell-embryo cDNA"  
 /dev\_stage="four-cell-embryo"  
 /dev\_stage="four-cell-embryo"  
 BASE COUNT 187 a 86 c 129 g 182 t 5 others  
 ORIGIN  
 Query Match 7.7%; Score 19; DB 44; Length 589;  
 Best Local Similarity 100.0%; Pred.No.1.8; 0; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;  
 Oy 202 TTTGTTTAAACCAAGAAA 220  
 ||||||||||||||||||  
 Db 484 TTTGTTTAAACCAAGAAA 502

Search completed: August 6, 1999, 09:22:59  
 Job time: 6250 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:14:23 ; Search time 3264.21 Seconds  
(without alignments)  
241.626 Million cell updates/sec

Title: US-09-049-696-14

Perfect score: 248  
Sequence: 1 ACCTGAGCGGGAATTTCAC.....TTGAAATGGCACAGATCTT 248

Scoring table: IDENTITY\_MDC

Searched: 679419 segs, 1590154680 residues

Database :

GenBdb1: \*  
1: gb\_da1: \*  
2: gb\_da2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_sts: \*  
14: gb\_sy: \*  
15: gb\_un: \*  
16: gb\_vl: \*  
17: em\_fun: \*  
18: em\_hlg: \*  
19: em\_hum1: \*  
20: em\_hum2: \*  
21: em\_hum3: \*  
22: em\_hum4: \*  
23: em\_om: \*  
24: em\_or: \*  
25: em\_ov: \*  
26: em\_pat: \*  
27: em\_ph: \*  
28: em\_pl: \*  
29: em\_ro: \*  
30: em\_sts: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_vl: \*  
34: gb\_hlg1: \*  
35: gb\_hlg2: \*  
36: gb\_in1: \*  
37: gb\_in2: \*  
38: em\_da1: \*  
39: em\_da2: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	248	100.0	878	5	195746	195746 Sequence 8

2	248	100.0	3311	11	AF039400	AF039400 Homo sapi
3	248	100.0	2826	42	AF127036	AF127036 Homo sapi
4	169	68.1	35278	11	AF039401	AF039401 Homo sapi
5	149.4	60.2	2937	12	AB017156	AB017156 Mus muscu
6	94.6	38.1	2984	3	BT036445	BT036445 Bos taurus
7	88	35.5	3471	12	AF052746	AF052746 Mus muscu
8	87.4	35.2	3317	13	AF001261	AF001261 Bos tauru
9	87	35.1	3022	12	AF047838	AF047838 Mus muscu
10	82	33.1	3415	42	AF043976	AF043976 Homo sapi
11	60	24.2	3604	9	AB026833	AB026833 Homo sapi
12	36.6	14.8	40738	8	YSC19470	Y17246 Saccharomyc
13	36.6	14.8	35583	36	CEC297	292872 Caenorhabdi
14	36.4	14.7	14320	37	CELPAR2	U00025 Caenorhabdi
15	36.2	14.6	110000	34	CEY71A12.2	Continuation (3 of
16	36.2	14.6	53202	8	AF072271	Continuation (4 of
17	35.6	14.4	13914	8	AF072271	AF072271 Lactuca s
18	35.2	14.2	4325	12	MMU66472	U66472 Mus musculu
19	35	14.1	236120	17	AF063866	AF063866 Melanoplu
20	34.6	14.0	110000	34	CEY37H9_0	AF022896 Caenorhab
21	34.6	14.0	30782	36	CEP33A8	Z81525 Caenorhabdi
22	34.6	14.0	38122	36	CELF45E1	U28732 Caenorhabdi
23	34.4	13.9	291821	34	CEY48E1	Z92856 Caenorhabdi
24	34.4	13.9	213763	35	AC007367	AC007367 Homo sapi
25	34.4	13.9	84704	36	CEY48E1B	Z93393 Caenorhabdi
26	34.2	13.8	35365	36	CELP37D9	U49829 Caenorhabdi
27	34	13.7	291692	34	CEY59A8	Z98870 Caenorhabdi
28	34	13.7	172568	42	AC007016	AC007016 Homo sapi
29	33.8	13.6	74717	37	AC004563	AC004563 Drosophi1
30	33.6	13.5	1320	3	CEY27808	U27808 Canis fami1
31	33.6	13.5	1320	3	DOGDC12G	L42513 Canis fami1
32	33.6	13.5	96475	7	ATP17M5	AL035678 Arabidops
33	33.6	13.5	98124	8	ATT16L1	AL031394 Arabidops
34	33.6	13.5	151862	34	HS376H23	AL031078 Homo sept
35	33.4	13.5	2911	1	SRTLYBA	X73140 S.hydoyent
36	33.4	13.5	337565	34	CEY47D3	Z98865 Caenorhabdi
37	33.4	13.5	296699	34	CEY48A6	Z92854 Caenorhabdi
38	33.2	13.4	131353	10	HS508I15	AL021707 Human DNA
39	33.2	13.4	132090	12	AC005402	AC005402 Mus muscu
40	33	13.3	1984	5	I18370	I18370 Sequence 25
41	33	13.3	1985	5	I21357	I21357 Sequence 25
42	33	13.3	1985	5	I40346	I40346 Sequence 25
43	33	13.3	1984	7	GBU04267	U04267 Gossypium b
44	33	13.3	122469	10	HS905G11	AL035045 Human DNA
45	33	13.3	155880	34	AC006739	AC006739 Caenorhab

## ALIGNMENTS

RESULT	1	LOCUS	195746	878 bp	DNA	PAT	17-JUL-1998
DEFINITION	Sequence 8 from patent US 5733748.						
ACCESSION	195746						
NUID	93940216						
VERSION	195746.1	GI:3940216					
KEYWORDS	Unknown.						
SOURCE	Unknown.						
ORGANISM	Unclassified.						
REFERENCE	1 (bases 1 to 878)						
AUTHORS	Yu, G. and Rosen, C.						
TITLE	Colon specific genes and proteins						
JOURNAL	Patent: US 5733748-A 8 31-MAR-1998;						
FEATURES	location/Qualifiers						
source	1..878						
BASE COUNT	257 a 179 c 188 g 241 t						
ORIGIN	100.0%; Score 248; DB 5; Length 878;						
Query Match	100.0%; Score 248; DB 5; Length 878;						
Best Local Similarity	100.0%; Pred. No. 1.7e-57;						

Matches	248;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	ACCTGAAGGCGGAATTCACGGGGCGCAGTCTCATTAATCTGACTGGACAGCTCTCTGGGG	60						
Db	306	ACCTGAAGGCGGAATTCACGGGGCGCAGTCTCATTAATCTGACTGGACAGCTCTCTGGGG	365						
QY	61	ATGATTATGACCATCGAAGACAGCTCACAAGTATTCATTCGAATTAAGTACAGTATCTTG	120						
Db	366	ATGATTATGACCATCGAAGACAGCTCACAAGTATTCATTCGAATTAAGTACAGTATCTTG	425						
QY	121	ATTCAGAGACAATTTAATGAATCTCTTCAAGTAATTAATTAATTAATTAATTAATTAATTA	180						
Db	426	ATTCAGAGACAATTTAATGAATCTCTTCAAGTAATTAATTAATTAATTAATTAATTAATTA	485						
QY	181	AAGCCAACTCTGAGCAAGTCTTTTGTGTTTAAACCAAGAAACATTACTTTGAAAATGGGA	240						
Db	486	AAGCCAACTCTGAGCAAGTCTTTTGTGTTTAAACCAAGAAACATTACTTTGAAAATGGGA	545						
QY	241	CAGATCTT 248							
Db	546	CAGATCTT 553							
RESULT	2								
LOCUS	AF039400	3311 bp	mRNA	PRI	15-DEC-1998				
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hclcal1) mRNA,								
ACCESSION	AF039400								
NID	94009457								
VERSION	AF039400.1	GI:4009457							
KEYWORDS	human.								
SOURCE	Homo sapiens								
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;								
REFERENCE	Eutheria; Primates; Catarrhini; Homnidae; Homo.								
AUTHORS	Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.								
TITLE	Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins								
JOURNAL	Genomics 54 (2), 200-214 (1998)								
MEDLINE	2 (bases 1 to 3311)								
REFERENCE	Gruber,A.D., Elble,R. and Pauli,B.U.								
AUTHORS	Direct Submission								
TITLE	Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,								
JOURNAL	Cornell University, Ithaca, NY 14853, USA								
FEATURES	Location/Qualifiers								
source	1..3311								
gene	/organism="Homo sapiens"								
CDs	/db_xref="taxon:9606"								
	/chromosome="1"								
	/map="1p22-p31"								
	/tissue_type="small intestine"								
	1..3311								
	/gene="hclcal1"								
	352..3096								
	/gene="hclcal1"								
	/note="transmembrane glycoprotein"								
	/codon_start=1								
	/protein_id="AAC95428.1"								
	/db_xref="PID:94009458"								
	/db_xref="GI:4009458"								
	/translation="MGPKRSVFIIILHLEGAISNSLIQLNNNGYICIVYADPNP								
	EBEITLQIQIKDWTQASLYLEFATGKRFFYFNVAI,IPETKRTADYVRKLETKNA								
	DVLVASEPSPGNDEPYTEOMGNGCGEGRILTPDPIAGKLAAYGOGAKAFVEMAH								
	LMKQVDEVDNNKEKYLISNGRTQAVRCSAGITGNVYKCGGSCYTKRCPFNKTYI								
	YKGGCFVLSQKOTERKASIMPAQHDSTVEPTCTONNHKSGAPNQNOKNLSRWVEYI								
	RTSEDEKTPTPMTQPNPNTSLDIQIGRTIVCIYIDKSGMATGNLRLNINOGGLFL								
	LOTVEDEGSVNGAVTDSAAHWSELTIQINSGSDRDTLAKRLPAASGCTSTCSLRBA								

BASE COUNT	1028 a	692 c	742 g	849 t
ORIGIN	<p>           FTVTRKRYPPDSDSEYIULNDGENTISGCFNFGOSGALIHFVLAAGPAADELELEL            MTGILQIYVYASDVONNGLIDAFGALSSGNGAYSQSILOEKGLTIONSOMNGTIVY            DSTYGRKTLFLITWTQPPQILMDPSGKOGGEYVDRKTKAHILQIDGIAKVGWTWK            SLOASQSLTLIVTSRASNAITLPPPIVTSKTRKDSKPSPLVYANRQGASFLTRA            SVTLAISBNGKIVTLELIDNGAGADATDDVYSRYFTYDINGRSVKVRALGVNA            AARRVRIPOOSGALIPGWENDEIOWMNPPEIKMDVOHQVCFSTSSGSEFVAS            DVPNRAPIDLEFPFQIOTIDUKAEIIFHGGSLINLMTAPGDDYDHGTAKHYIIRISIDLD            LDPKFNESLOVNTALIPKRNASEEYFLKRPENITPENOTDLFIALOKVDRVDKSELE            SNIRVSLFIEPIOTPETPSPDETSAPCFPIHINSTIIPGIIHLIKIMKMIQELDLSTA            "         </p>			
Query Match	100.0%; Score 248; DB 11; Length 3311;			
Best Local Similarity	100.0%; Pred. No. 1,6e-57;			
Matches 248; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
OY	<p>           1 ACTGGAAGCGGGAATTCACGGGGGACATCCATTAATCTGACTTGACAGCTCTGGGG 60                               DB 2624 ACCGGAAGGGGAAATTCACGGGGGACATCCATTAATCTGACTTGACAGCTCTGGGG 2683                               OY 61 ATGATTATGACATGGAACAGCTCAAGATATATCATTCGATCGAATAGTACAAGATTCTTG 120                               DB 2684 ATGATTATGACATGGAACAGCTCAAGATATATCATTCGATCGAATAGTACAAGATTCTTG 2743                               OY 121 ATCTCAGACAAAGTTCAATGAATCTCTTCAAGTGAATATCTACTGCTCTATCCCAAGG 180            2744 ATCTCAGACAAAGTTCAATGAATCTCTTCAAGTGAATATCTACTGCTCTATCCCAAGG 2803                               OY 181 AAGCAACCTGTGAGGAAGCTTTTGTGTTAAACAGAAACATTACTTGTGAAATGCGA 240                               DB 2804 AAGCAACCTGTGAGGAAGCTTTTGTGTTAAACAGAAACATTACTTGTGAAATGCGA 2863                               OY 241 CAGATCTT 248                               DB 2864 CAGATCTT 2871         </p>			
RESULT 3	<p>           AF127036 2826 bp mRNA PRI 16-Apr-1999            LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL1)            DEFINITION mRNA, complete cds.            AF127036            94585468            AF127036.1 GI:4585468            .            human.            ORGANISM Homo sapiens            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;            Eutheria; Primates; Catarrhini; Homiidae; Homo.            REFERENCE 1 (bases 1 to 2826)            AUTHORS Agnel,M., Verman,T., and Cuioussou,J.-M.            TITLE Cloning of three human homologs of bovine epithelial chloride            channel            JOURNAL Unpublished            2 (bases 1 to 2826)            AUTHORS Agnel,M. and Cuioussou,J.-M.            TITLE Direct Submission            JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des            Carrieres, Neuilly-Malmaison 92500, France            FEATURES            source            1..2826            /organism="Homo sapiens"            /db_xref="taxon:9606"            /tissue_type="small intestine; colon"            1..2826            /gene="CACCL1"            5..2749            /gene="CACCL1"            /note="bovine epithelial chloride channel homolog"            /codon_start=1         </p>			



```

exon      19783..20010
           /gene="hclcal"
exon      /number=8
           22253..22427
           /gene="hclcal"
exon      /number=9
           24522..24628
           /gene="hclcal"
exon      /number=10
           26644..26859
           /gene="hclcal"
exon      /number=11
           27447..27708
           /gene="hclcal"
exon      /number=12
           28766..28936
           /gene="hclcal"
exon      /number=13
           31840..32079
           /gene="hclcal"
exon      /number=14
           32919..33522
           /gene="hclcal"
exon      /number=15
           33522..33522
           /gene="hclcal"

BASE COUNT      11245 a      6870 c      7230 g      9933 t
ORIGIN

```

```

Query Match      68.1%; Score 169; DB 11; Length 35278;
Best Local Similarity 100.0%; Pred. No. 2.8e-36;
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      80 AGCTCAGAGTATATCTTGCATAGTACATATCTTGCATAGTACAGAGTCA 139
           |||
DB      32917 AGCTCAGAGTATATCTTGCATAGTACATATCTTGCATAGTACAGAGTCA 32976
           |||

QY      140 TGAATCTCTTCAAGTGAATGATCTGCTCTCATCCCAAGAGCCAACTCTGAGAGT 199
           |||
DB      32977 TGAATCTCTTCAAGTGAATGATCTGCTCTCATCCCAAGAGCCAACTCTGAGAGT 33036
           |||

QY      200 CTTTGTGTTAAACGAGAAACATTACTTTGAAATGGCAGATCTT 248
           |||
DB      33037 CTTTGTGTTAAACGAGAAACATTACTTTGAAATGGCAGATCTT 33085
           |||

```

```

RESULT      5
LOCUS      AB017156      2937 bp      mRNA
DEFINITION Mus musculus gob-5 mRNA, complete cds.
ACCESSION      AB017156
NID      93721911
VERSION      AB017156.1 GI:3721911
KEYWORDS      GOB-5
SOURCE      Mus musculus adult intestine goblet cell cDNA to mRNA.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE      1 (sites)
AUTHORS      Komiyama, T., Tanigawa, Y. and Hirohashi, S.
TITLE      Cloning and identification of the gene gob-5, which is expressed in
              intestinal goblet cells in mice
JOURNAL      Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)
MEDLINE      99160866
JOURNAL      2 (bases 1 to 2937)
REFERENCE      Komiyama, T., Tanigawa, Y. and Hirohashi, S.
AUTHORS      Submitted (26-AUG-1998) to the DBJ/EMBL/GenBank databases. Tooru
              Komiyama, Eranio, Jst, Hirohashi Cell Configuration Project; 5-9-9,
              Tokoda, Tsukuba, Ibaraki 300-2635, Japan
              (E-mail: tkom@hccp.jst.go.jp; Tel: 81-298-47-7563,
              Fax: 81-298-47-5226)
FEATURES
SOURCE      Location/Qualifiers
              1..2937

```

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="goblet cell"
/dev_stage="adult"
/tissue_type="intestine"
15..2756
/gene="gob-5"
15..2756
/gene="gob-5"
/codon_start=1
/protein_id="BA33743.1"
/db_xref="PIR:d1034712"
/db_xref="PIR:g3721912"
/db_xref="GI:3721912"
/translation="MESLKSPEVLLIHLLEGVLSSELIOLNNGEGYIADHDVP
EDEALIOHKDMVTOASPYLEFATGKRFYKNAIILPESWKKAPPTPRKLETFNA
DVLVSTSPLENDPEYTEHIGACGCKGIRHLHPDLGKGLTQYGPQDPTFHEWAH
FRMGVENEYNNDEKFLSKGPQAVRCSAAITGNQVRCQSGCLTNGCVLDRTG
LYKNCVFPDPDHPQNEKASIMFNONINSVVEFCEKHNHNOEAENDONRCNLSTVE
IOESDFKQTTPMTAOPAPPFESLQIGORIVCLVLDKSGSMLNDRNLNMSVLE
LQIVEGQSWGNTTSDQVONGLVDPAFALSSGNAALHOSHOLESRGVNLONNOMNSVY
AFTVTKRRYPIDGSEIVLTDGEDNTISSCDIYKOSGALIHITVALGPAAKELEQIS
KMTGGLQYSSDQVONGLVDPAFALSSGNAALHOSHOLESRGVNLONNOMNSVY
VDSYKQDLFLITWTHTPTIFIMDSGVQKQFILDPTTKYAVLYVPGTAVGEWK
YSIOASQTLTLVTSRAASATLPPIVTPVNNKNTGFSPVTVYSIOGASPLR
ASVTLAISVNGKTVTLLELDNGAGADATKNDGVSRFTAFDNGRSKRYNALGV
TSDROBAAPKPKRAYIDGWIEDGEVRNMPPEPETSIVODQQLCFSTSSGGSFVYN
VPAAPRIPDLFPQOITDLKASIOGONVNLMTAPGDYDHGRASNYIIMSTSYD
LNDHRTSLQVNTTGLIKKASSEIEFEFELGNTFGNGTDIFAIQAVKSNLSSEI
SNARVSVPFPAOEPPIDESTPPEDPDISINSTITPGIHLVKIMKMKLGEQVTLGLH"

```

```

BASE COUNT      860 a      718 c      693 g      666 t
ORIGIN

```

```

Query Match      60.2%; Score 149.4; DB 12; Length 2937;
Best Local Similarity 75.3%; Pred. No. 6.4e-31;
Matches 186; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```

```

QY      1 ACCGTAAGGCGGAATTTACGGGGGCACTCATTAATCTGACTTGGACAGCTCTGGGG 60
           |||
DB      2287 ACCGTAAGGCGGACATCCAGAGGCGGACGACTGGTAATCTGAGCTGAGCGCTCTGGGG 2346
           |||

QY      61 ATGATTATGACCATGAGAACGCTCACAAGTATATCATTCGATTAAGTACAAATTTCTTG 120
           |||
DB      2347 ATGATTATGACCATGAGAACGCTCACAAGTATATCATTCGATTAAGTACAAATTTCTTG 2406
           |||

QY      121 ATCTCAGACCAAGTTCAATGATCTCTTCAAGTGAATACTACTGCTCTCATCCCAAG 180
           |||
DB      2407 ATCTCAGGACCACTTCAACACCTCAGTCCAGTGAACACACGCGCTTATCCCAAG 2466
           |||

QY      181 AAGCAACTCTGAGGAATCTTTTGTTAACGAGAAACATTACTTTGAAATGGCA 240
           |||
DB      2467 AAGCAAGCTCTGAGGAATCTTTGAGTTGAACTGGGAGGCAACACTTTGAAATGGCA 2526
           |||

QY      241 CAGATCT 247
           |||
DB      2527 CAGATAT 2533
           |||

```

```

RESULT      6
LOCUS      BTU36445      2984 bp      mRNA
DEFINITION Bos taurus calcium-activated chloride channel mRNA, complete cds.
ACCESSION      U36445
NID      9184065
VERSION      U36445.1 GI:1184065
KEYWORDS
SOURCE      cow.
ORGANISM      Bos taurus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
              1 (bases 1 to 2984)

```



AUTHORS Cunnigham,S.A., Awayda,M.S., Bubbien,J.R., Ismailov,I.I., Arrate,M.P., Bertley,B.K., Benos,D.J. and Fuller,C.M.  
TITLE Cloning of an epithelial chloride channel from bovine trachea  
JOURNAL J. Biol. Chem. 270 (52), 31016-31026 (1995)  
MEDLINE 96125078  
REFERENCE 2 (bases 1 to 2984)  
AUTHORS Benos,D.J.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics, University of Alabama at Birmingham, BHSB 706, Birmingham, AL 35294, USA

FEATURES  
source Location/Qualifiers  
1..2984  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"  
/tissue\_type="trachea"  
19..2730  
/note="epithelial chloride channel"  
/codon\_start=1  
/product="calcium-activated chloride channel"  
/protein\_id="AAC48511.1"  
/db\_xref="PID:91184066"  
/db\_xref="GI:1184066"  
CDS  
/translation="MVPRLLVTLFLHLLEPGKSSMYNLINGYDGIYAINPSVE  
DEKLQNTKEMVTEASTYLFPHATKRRVYFRNYSILIPMTKSKSEYLIPKQSDYQAE  
VIVANPYLKHGDPTTLYQGRGEGKGYIHFPNLTNNLITSGRAVEVHAHL  
RMGIFDEYNGDPEYISRNRTIATRCSTHIGTVIVKCGGSCITRCRNDSTGL  
YEAKTFIPEKQTAIESIMFQSLHSTVEFTEKTHVEAENLONKKNCSITWDI  
MNSTDFONTSPTEKMPPTOPFSELSKQRYVCLVDRSGSSESDLEFRNOAEL  
FLIOIIEGSLGAVTIPDSVAEIRNNLTIKRDYNTENTALTPOEANGSTSICGKT  
AGFOAILOSOSTSSEITLLDGEDNEHSCIEVKOSGVITHIALGPSAKLEET  
LSDMTGGRFVANKDINGLTNLAFKSISSRSSTIQTQLESKALATIEKWNSTVP  
VDSTIGNDTFEVVWTIKRPEILLQDPKGRKXTSDFKEDKLINHSARLPGIAGETG  
TWTSILNNHSAPOILTVTTPRARSPTPTATAHSONTAHPSRVIYAYQVSG  
FLPVGINVTALIEEDGHVTLLEMDNGADTVKNGIYSRYFTDVRNGRYSILKY  
HAERANNARLSIRQPNKALYIPGIEYNGKILLNRPPEVYDANLIDGKANSYIIRI  
GGSFTVSGAPRGNHPSVLPNKIIDLKAKREDHILQSTAPANLIDGKANSYIIRI  
SKSFLLDKQDFDNATLVNTSSLKPRKAGSDENFEKPEPFRIENGTNYIAVOAINEA  
NLTSVSNIAQAIKRIIPMEDSVLPALGTRKISAINIAPALAMILISIV"

BASE COUNT 1027 a 583 c 561 g 813 t  
ORIGIN  
Query Match 38.1%; Score 94.6; DB 3; Length 2984;  
Best Local Similarity 65.3%; Pred. No. 3.9e-16;  
Matches 139; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 29 TCTCATTATCTGACTTGACAGCTCTCTGGGATGATATAGACCAAGCAAGCTCACAA 88  
DB 2337 TCACATTCACCTTATGACAGCGCCCTCGATGCTCAGATAAAGCAAGCCACAG 2396  
QY 89 GATATCATTCGAAATAGACAGATGATCTTGATCTCAGAGCAAGTAATGATCTCT 148  
DB 2397 CACACATTAAGAATAAGAGAGTTCTCTGATCTCCAAAAGATTTTGACATCTAC 2456  
QY 149 TCAAGTGAATCTACTGCTCTCATCCCAAGGAGCAACTCTGAGGAAGTCTTTTGT 208  
DB 2457 TTAGAGTACTCTCAGCTAAACCTAAGGAGCGGCTCAGATGAAAAATTTGAATT 2516  
QY 209 TAAACCAAGAAACATCTTTTGAAGAAATGGCAC 241  
DB 2517 TAAACCAAGAACTTTAGATAAGAAATGGCAC 2549  
RESULT 7  
AF052746 3471 bp mRNA ROD 08-APR-1999  
LOCUS AF052746 Mus musculus chloride channel cacc mRNA, complete cds.  
DEFINITION AF052746  
NID 93560546  
VERSION AF052746.1 GI:3560546  
KEYWORDS

SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 3471)  
TITLE Characterization of a murine gene homologous to the bovine Cacc  
JOURNAL Romo,L., Musante,L., Cintl,R., Serl,M., Moran,O., Zegarra-Moran,O.  
MEDLINE 99173882  
REFERENCE 2 (bases 1 to 3471)  
AUTHORS Romo,L., Musante,L., Cintl,R., Serl,M., Moran,O., Zegarra-Moran,O.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto  
Gianluca Gaslini, Genova I-16148, Italy  
FEATURES  
source Location/Qualifiers  
1..3471  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/db\_xref="dbEST:AA245677"  
47..2752  
/codon\_start=1  
/product="chloride channel cacc"  
/protein\_id="AAC35003.1"  
/db\_xref="PID:93560547"  
/db\_xref="GI:3560547"  
/translation="MVPEGLVLEFLTLHLNTESMYNLNSNGYGVIAINPSVE  
DEKLQNTKEMVTEASTYLFPHATKRRVYFRNYSILIPMTKSKSEYLIPKQSDYQAE  
VIVANPYLKHGDPTTLYQGRGEGKGYIHFPNLTNNLITSGRAVEVHAHL  
RMGIFDEYNGDPEYISRNRTIATRCSTHIGTVIVKCGGSCITRCRNDSTGL  
YEAKTFIPEKQTAIESIMFQSLHSTVEFTEKTHVEAENLONKKNCSITWDI  
MNSTDFONTSPTEKMPPTOPFSELSKQRYVCLVDRSGSSESDLEFRNOAEL  
FLIOIIEGSLGAVTIPDSVAEIRNNLTIKRDYNTENTALTPOEANGSTSICGKT  
AGFOAILOSOSTSSEITLLDGEDNEHSCIEVKOSGVITHIALGPSAKLEET  
LSDMTGGRFVANKDINGLTNLAFKSISSRSSTIQTQLESKALATIEKWNSTVP  
VDSTIGNDTFEVVWTIKRPEILLQDPKGRKXTSDFKEDKLINHSARLPGIAGETG  
TWTSILNNHSAPOILTVTTPRARSPTPTATAHSONTAHPSRVIYAYQVSG  
FLPVGINVTALIEEDGHVTLLEMDNGADTVKNGIYSRYFTDVRNGRYSILKY  
HAERANNARLSIRQPNKALYIPGIEYNGKILLNRPPEVYDANLIDGKANSYIIRI  
GGSFTVSGAPRGNHPSVLPNKIIDLKAKREDHILQSTAPANLIDGKANSYIIRI  
SKSFLLDKQDFDNATLVNTSSLKPRKAGSDENFEKPEPFRIENGTNYIAVOAINEA  
NLTSVSNIAQAIKRIIPMEDSVLPALGTRKISAINIAPALAMILISIV"

BASE COUNT 1096 a 777 c 736 g 862 t  
ORIGIN  
Query Match 35.5%; Score 88; DB 12; Length 3471;  
Best Local Similarity 63.0%; Pred. No. 2.4e-14;  
Matches 136; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 33 ATTAATCTGACTTGACAGCTCTCTGGGATGATATAGACCAAGCAAGCTCACAA 92  
DB 2363 ATTACCTTACATGACAGCGCCCTCGAAGTTCGCAAAATGGAAGCAATGATAC 2422  
QY 93 ATCAATCGAATAGTACAGATATCTTGATCTCAGAGCAAGTCAATGATCTTTCAA 152  
DB 2423 ATCAATCGAATAGTACAGATATCTTGATCTCAGAGCAAGTCAATGATCTTTCAA 152  
QY 153 GTGAATACACAGCTCTCTCCCAAGGAGCAAGCAAGTCTGAGGAAGTCTTTTGA 212  
DB 2483 GTGAATACACAGCTCTCTCCCAAGGAGCTCTCAAAAAGCAATTAATTAACAA 2542  
QY 213 CCAGAAACATTAATTAATTAATGCAATGCAATGCAATGCAATGCAATGCAAT 248  
DB 2543 CCAGAAACATTAATTAATTAATGCAATGCAATGCAATGCAATGCAATGCAAT 2578  
RESULT 8  
AF001261 3317 bp mRNA MAM 19-NOV-1997  
LOCUS AF001261 Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA.  
DEFINITION



DB 2334 ATTCACCTTACATGAGAGGCCCTGGCAAGCTCTGACAAATGAGAGACATAGATAC 2393  
QY 93 ATCATTCGAATAGTACAGATATCTTGATCTCAGAGACAGATTCATGATCTCTCAA 152  
DB 2394 ATCATTCGAATAGTACAGATATCTTGATCTCAGAGACAGATTCATGATCTCTCAA 2453  
QY 153 GTGAATACACTGCTCTCATCCCAAGAGACCACTCTGAGAGAGCTTTTGTTTAA 212  
DB 2454 GTGAATGCTTCAGCTGATCTGATCTCAAGAGAGCTCAAGAGCAATTTAAATTCAAA 2513  
QY 213 CCAGAAACATCTACTTTGAAATGGACAGATCT 247  
DB 2514 CCAGAAACATTTTAAATAGCAATGGCATCCAGCT 2548  
RESULT 10  
AF043976 3415 bp mRNA PRI 07-APR-1999  
LOCUS Homo sapiens CLCA homolog (hCLCA3) mRNA, complete cds.  
ACCESSION AF043976  
NID 94572288  
VERSION AF043976.1 GI:4572288  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 3415)  
JOURNAL Gruber, A.D., Elble, R.C. and Pauli, B.U.  
Molecular cloning and biochemical characterization of a truncated,  
secreted member of the human family of Ca2+-activated Cl- channels  
Biochim. Biophys. Acta 1444 (3), 418-423 (1999)  
JOURNAL 2 (bases 1 to 3415)  
MEDLINE 9916715  
AUTHORS Gruber, A.D., Elble, R.C. and Pauli, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1. .3415  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1. .3415  
/gene="hCLCA3"  
19. .807  
/gene="hCLCA3"  
/codon\_start=1  
/product="CLCA homolog"  
/protein\_id="AAD23734.1"  
/db\_xref="pid:94572289"  
/db\_xref="gi:4572289"  
/translation="MVFLKVLFLSLLSPLYKSLVLTNNNGYDGIYAIINPSVPE  
DEKLQNIKEWTEASTLHFAIKORAFRANVSIIPMYKSESLIPKQYDQAD  
VIVADLYLKGDPTLQYGGCGGGOYIHFNPFLTNLNLATYGRGVFVGHANL  
RMGVDEVNDQPEYISRNTEATRCSTRIVVAVLNDCCAGASCIARFRDSDTGL  
YEAKTFIPKSKTAKESIVFNONIDSTEFCTEXTHNKEAPNL"  
BASE COUNT 1149 a 625 c 630 g 1011 t  
ORIGIN  
Query Match 33.1%; Score 82; DB 42; Length 3415;  
Best Local Similarity 61.0%; Pred. No. 9.8e-13;  
Matches 133; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

DB 2506 GCTGCTTATTAATATCTTGCTGATACCTTAACGACACTGGTTCAGTAGAAGTTT 2565  
QY 204 TTGTTTAAACAGAAAACATTTACTTTTGAATAAGCAGC 241  
DB 2566 GATTTTAAACAGAACCTTCTAAATAGAAATGATG 2603  
RESULT 11  
AB026833  
LOCUS AB026833 3604 bp mRNA PRI 26-MAY-1999  
DEFINITION Homo sapiens mRNA for chloride channel protein, complete cds.  
ACCESSION AB026833  
NID 94867600  
VERSION AB026833.1 GI:4887600  
KEYWORDS chloride channel protein.  
SOURCE Homo sapiens adult corneal epithelium cDNA to mRNA,  
clone.lib.lambda zap clone.lambda 13802.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (sites)  
JOURNAL Itoh, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.  
Isolation and characterization of chloride channel in human corneal  
epithelium  
Unpublished (1999)  
2 (bases 1 to 3604)  
AUTHORS Itoh, R., Kawamoto, S. and Okubo, K.  
TITLE Direct Submission  
JOURNAL Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases. Relko  
Itoh, Institute for Molecular and Cellular Biology, Osaka  
University, Yamada-oka 1-3, Suita, Osaka 565-0871, Japan  
(E-mail:relko@imcb.osaka-u.ac.jp, Tel:81-6-6879-7992,  
Fax:81-6-6877-1922)  
FEATURES  
source  
1. .3604  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="lambda 13802"  
/clone.lib="lambda zap"  
/dev\_stage="adult"  
/map="1p32"  
/tissue\_type="corneal epithelium"  
42. .2873  
/codon\_start=1  
/product="chloride channel protein"  
/protein\_id="BA27810.1"  
/db\_xref="pid:d1041578"  
/db\_xref="pid:94867601"  
/db\_xref="gi:4887601"  
/translation="MTORSIGAPICNLKFTVLVALSSSELPGLAGVOLQDNGYNGLL  
IAINQVENVONLISNIKEMITEASEFYFNATKRFRVFNRIKILIPATWKKANNKIK  
QSEYKAVAYITDVGAGDDPPTLVQAGDKVCLVDSVSMADRLDQQA  
VFYEMALRNGVDEYINDPPTLVQAGDKVCLVDSVSMADRLDQQA  
SKLREGECTFLYNSTONATASIMWQSSSVVERCNASTHQAENADLNQKSLASW  
DVIDTADFHSHSPMNGTELPPTFLVQAGDKVCLVDSVSMADRLDQQA  
AEFTYLMQVEIHEITFVIGIASPDSKEIRAOQINSNDKRLVSLPTVSAKDISI  
CSGLKGEFEVEKLNKAGYSVMILVTSQDRLGNCPLVLSGSLTSHLSALGSSA  
PNLELSRLTGKIFEPDINSMDIFRISGRTDIFQOHQLESDEANYPH  
QKNTVYDNTVNDTMTLVQWASGPPEILFPDQKKYTNNTITLTERTSMT  
QPTAKPGWITTLNTHSLQALKYTVTSRANSNAPPAIVEADVERDSLHFPEVM  
YANWQGGYPTLNTAVTAVEPDTLRLDGAADVIKNDGITSRYFSEFAN  
GYSLSKLVNHSPSISTPAHSIPGSHAVYGVYANGNIQNNAPRKSQVREERKQW  
FSRVSQGSFVLGVPAGPHDVPCKIIDEAVKVEEELTSLTAGDEPDQOAT  
SEYTRMSKSLQNDENNALIVTSKRNPOAGIREFTSPQISTGPEPHONGET  
HSERIVAYIAMDNRNSQASVNSIAQPLETIPRNSDVPARDVILIKGVLTAMGLIG  
ITCLIVTHTLSKRRADKENGTKLL"  
BASE COUNT 1113 a 740 c 732 g 1017 t 2 others  
ORIGIN  
Query Match 24.2%; Score 60; DB 9; Length 3604;



```

/codon_start=1
/db_xref="SGD:S0004160"
/product="Apsip: clathrin coat assembly protein complex,
small subunit"
/protein_id="AAB67468.1"
/db_xref="PID:9577208"
/db_xref="GI:577208"
/translation="MTOLKYLILSVROGRIRLKKWTAMSGEKATIVDLPPTILAR
KPMCNITIERNDKVVYKRYKASLFIYGMTPDVNDLLEITLHREVTEDIDYFNAV
EDIDIFNRSKVDILNEMIMCDGSIASSSRKREVLHHTVMDTWESNDLEIRVLS"
2897..3286
/gene="YLR171W"
2897..3286
/gene="YLR171W"
/codon_start=1
/product="Ylr171wp"
/protein_id="AAB67478.1"
/db_xref="PID:9234098"
/db_xref="GI:234098"
/translation="MSSQTLPRKYSIVSNRWIISKVSSSLTSGVPIPIKRLAY
LTTTMSLSMMLHIGFLANITGVKSFITFASPADIATVTHFFNLIFPLETSNKIF
NCVILCTGVSVINLQDRSCSWLKL"
complement(3424..4326)
/gene="DPH5"
/db_xref="SGD:S0004162"
complement(3426)
/gene="DPH5"
/translation="S-adenosylmethionine (AdoMet)-dependent
methyltransferase of diphthamide biosynthesis"
/codon_start=1
/db_xref="SGD:S0004162"
/product="Dph5p: diphthine synthase"
/protein_id="AAB67469.1"
/db_xref="PID:9577209"
/db_xref="GI:577209"
/translation="MLYLIGLSYKSDITVRGLFAIKKRSRYLEHYTSILMAASOE
LESYVGEKLIADRELVEFGKQIINMNDKEVAIVYGDPPGATVYDPRATVIRARE
AIPVEITINNASVNAAGAGCLOXNFCQSVWFPPDMNRPRDPMKYNKINRGIGHT
LVLDIKVKEQSIENMARGRLITEPRINSIACCCQLEITEKRTKATPDTIGVA
ISRLGSSSQSFKSGTISELANVDSGPELSLVLIGRQCHELEYLELLEFADKKEFGK
DVANDVEYKPPAAMPPTEDDSE"
4585..6411
/gene="YLR173W"
4585..6411
/gene="YLR173W"
/codon_start=1
/evidence="not experimental"
/product="Ylr173wp"
/protein_id="AAB67465.1"
/db_xref="PID:9577205"
/db_xref="GI:577205"
/translation="MHTNSPLRANODLETOPILRPNTESOLLNDEVIRANETILI
KSRWSIKLILILGIVLSPFGVSIYQIRGHVPTDIEKNLQVNFKLEFOL
DKWKMGGDLNNDGKLYOVSIHQIOWEDYDGPPTENDSROQRDMRYNEKVL
KILCIDLNVTTFEDGLVFKKRGVGEPICFNLAHQINNLQKIKIVKSPINELV
KVLKFWKDFESLNTKSNLDTIKRKRGTRNLKLNDEILDMQDITWETSATP
LRQIMNIDGISTLQGTPLDSSSDGFHADRLNPITILGGVWLHLPPGSIPIFWME
IKIPDNGEPALNIPITLSCFNEPIINLHDKNDIVYCLQNEIEGPLEDELYOCFONS
LTPMSQIVAVLNQNETVFAARGHVLIEDGIDNLSLIPADMLEDIOEASLFIITNA
TFNSELIOEFOINDOLNMAARKSLIVGTGFGPDLSPFETHODRVRIDIRGOI
DIYHNDINFLNLPKRWIMSSSHLHDEDDGNOMKLOPDLNDEHVNNSLELTNTL
NELIFGTFVIHFNATIDASLTITLALGPWVLITGLAGSDILYT"
6755..7993
/gene="IDP2"
6755..7993
/gene="IDP2"
/translation="cytosolic form of NADP-dependent
isocitrate dehydrogenase"
/codon_start=1
/evidence="not experimental"
/product="Idp2p: isocitrate dehydrogenase"
/protein_id="AAB67464.1"
/db_xref="PID:9577204"

```

```

/db_xref="GI:577204"
/translation="MKRIKVPYEMDDEQTRIMHIRKIVLPYIDVLTQYIDL
SVEYKQINDQYVDSATATLKYGVAAVKCATITLPDAREEELHKKMKSPPGTLINKI

Query Match      14.8%  Score 36.8; DB 8; Length 40738;
Best local Similarity 50.6%; Pred. No. 1.3;
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 58 GGGATGATTATGACCATGAGACGCTCACAGTATATCATTCGAATAGTACAGATTTC 117
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36125 GAGGTAAAAAGCAAAAGCAAAATCTCAAAATCCATCCGAGACGACCAAGATTA 36066
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 118 TTGATCTCAGACCAAGTTCATGAATCTCTTCAAGGATTAATCTACTGCTCATCCAA 177
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36065 GACAAAAGAAAGCAACTTATTCGACTTTCATATTAAGTGGAAAAACGTCAAAAAGAC 36006
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36005 AATGTGCTAAGTTGAGACACGCGTTTGTTCAGACGACAGACACATTTTGAA 35950
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
CEC29F7      35583 bp      DNA      INV      23-NOV-1998
LOCUS      Caenorhabditis elegans cosmid C29F7, complete sequence.
DEFINITION      292827
ACCESSION      92266438
NID      292827.1 GI:2266438
VERSION      HTG.
KEYWORDS      Caenorhabditis elegans.
SOURCE      Caenorhabditis elegans.
ORGANISM      Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
      Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
      1 (bases 1 to 35583)
      McMurtry,A.
      Direct Submission
      Submitted (09-MAR-1997) Louis, MO 63110, USA. E-mail:
      jesse@sanger.ac.uk or iwenematode.wustl.edu
      2 (bases 1 to 35583)
      Wilson,R., Almscough,R., Anderson,K., Baynes,C., Berks,M.,
      Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
      Coulson,A., Craxton,A., Dear,S., Du,Z., Durbin,R., Favello,A.,
      Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
      Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
      Latreille,P., Lightning,J., Lloyd,C., McMurtry,A., Mortimore,B.,
      O'Callaghan,M., Parsons,J., Percy,C., Rikken,L., Roopra,A.,
      Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E.,
      Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
      Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
      Wilkinson-Sproat,J. and Woldman,P.
      2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
      elegans
      Nature 368 (6466), 32-38 (1994)
      94150718
      JOURNAL      Nature
      MEDLINE      94150718
      COMMENT      On Jul 19, 1997 this sequence version replaced gi:1877066.
      Coding sequences below are predicted from computer analysis, using
      predictions from GeneFinder (P. Green, U. Washington), and other
      available information.
      For a graphical representation of this sequence and its analysis
      see:
      http://webc.sanger.ac.uk/cgi-
      bin/display?db=wormbase&class=Sequence&object=C29F7
      Current sequence finishing criteria for the C. elegans genome
      sequencing consortium are that all bases are either sequenced
      unambiguously on both strands, or on a single strand with both a
      dye primer and dye terminator reaction, from distinct subclones.
      IMPORTANT: This sequence is NOT necessarily the entire insert of
      the specified clone. It may be shorter because we only sequence
      overlapping sections once, or longer because we arrange for a small
      overlap between neighbouring submissions.
      IMPORTANT: This sequence is not the entire insert of clone C29F7.
      It may be shorter because we only sequence overlapping sections

```





Db 103873 GATGTACCATGTGTAACTATTTTACACATCAAAACATTGAGTGAATTTG 103817  
Search completed: August 5, 1999, 23:14:38  
Job time: 7/61 sec

KTSHSVUGLLFNNVTNNVNFLLAKDSTYINPEVLLRMIDTWNNPEVYVMEGAEDG  
SGRCRLDTGMLISOPAHALMNNRNACNNFALADDDOLAEFCIOIATNLCCKPIHO  
GVREYVWGAERADSPAHADSIEDWKSHPAFKRLAVPRLSPADSAHALDYVREYM  
ORADREIIRKMEAEISRLAEQEARTEGAIISMPALPEYAPRNRYVSTWEITTELE  
FRSPNNQNRLEKGFDDVAEVVAARQOVESEEPLEVLQNGRVFDPGRGDY  
MVLITRYKTVNMEVDNREFSDNEAHEESLKEIYVERHVSRMIASTOLMNOAY  
VKEDDTYVTVPVASEKDYLPARKLARQRLCFPEEARHKTSMVAVAPLIESRSY  
TAITNDMEELKRRCKRSLLETDYLPVPAVSTGGKTAALDADAVRGANTITYL  
LSPHADVQKEFEEDRKRINTIKHYVFPVPEVHEHPRTISGEMTEKEKETPEQARE  
AALSRLDGVPEPKRRLIVQKEGRFDSODESCFAYGVDTYTAARKFQGNRRNDL  
ISAFIQDSIHVLAPEPTLRIRYHKRSCDMESIDTEDIARCLDSKKNVAAKQOLAK  
LLEHEK"  
13477..14316  
/gene="PAR2.1"  
join(13477..13569,13622..13686,13962..14316)  
/gene="PAR2.1"  
/note="similar to E. coli plasmid RK2 gene for p116 and  
single stranded binding protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AA50616.1"  
/db\_xref="GI:458885"  
/db\_xref="PID:943883"  
/translation="MLRSLSTISKSTVRCMSLTSKMAEOPSKOEVDLREKPOHN  
PEORRHAYSVNKLGVGLDPLKTRNGKPYLIFNITNSYFKQDDGTTLDQTER  
HAYSVFGQAEILSKTIKKSHLMVQGRHLYSGQKDEQGNRTQNTYIIAQTVQPLA  
RAARENPOH"  
BASE COUNT 4832 a 2441 c 2519 g 4528 t  
ORIGIN

Query Match 14.7%: Score 36.4; DB 37; Length 14320;  
Best Local Similarity 51.9%: Pred. No. 1.8;  
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 76 GAACAGCTCAGAGTATATCATCTGAATAGTACAGTATCTTGATCTCAGAGCAAGT 135  
Db 1612 GAAAGAGTTAGAAAAAATCCGCCAAAAAATTTTGTGGTCTCAAAAAATGC 1553  
QY 136 TCATGATGATCTCTCAAGTACATCTCTCATCCCAAGGAGCCAACTGTAGG 195  
Db 1552 TCAAAAAGTTTGTTCGAAATCTTGATTTTATACAAATATGATGAACATGTGC 1493  
QY 196 AAGTCTTTTGTTTAAACCGAAAAACATTACTTTTGA 233  
Db 1492 AAACATATTATTGTCACATGTTACATCTTTTGTGA 1455

RESULT 15  
CEY71A12\_2/c  
WPCOMMENT  
Sequence split into 4 fragments LOCUS CEY71A12 Accession AL021390

Fragment Name	Begin	End
CEY71A12_0	1	110000
CEY71A12_1	100001	210000
CEY71A12_2	200001	310000
CEY71A12_3	300001	353202

Continuation (3 of 4) of CEY71A12 from base 200001 (AL021390 Caenorhabditis elegans chrc

Query Match 14.6%: Score 36.2; DB 34; Length 110000;  
Best Local Similarity 50.3%: Pred. No. 1.8;  
Matches 89; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 61 ATGATTATGACCATGAGACCTCACAAGTATATCATTCGAAATAGTACAGTATCTTG 120  
Db 103993 ATATATTAAAAAACAACAAAGTTTGGGAAATCGTCAAAAAATTTATTTTGTGCG 103934  
QY 121 ATCTCAGAGACAGTTCATGATCTCTCAAGTGAATGATCTGCTCATCCCAAGG 180  
Db 103933 ATCTTCAAAAATCTTCAAAAAATTTTGTGCAAAATCTTGAAATTTTATACAAAATT 103874  
QY 141 AAGCCAACTCTGAGAGAGTCTTTTGTTTAAACGAAACATTACTTTGAAAATG 237



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:13:55 ; Search time 3264.21 Seconds

(without alignments)  
221.165 Million cell updates/sec

Title: US-09-049-696-13

Sequence: 1 GTGGGGCTCTGGGAGAGT.....AATGCTCCCATCACTGATCT 227

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_da1:\*  
2: gb\_da2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hlg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hlg1:\*  
35: gb\_hlg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_da1:\*  
39: em\_da2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pt4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	227	100.0	878	5	195746	Sequence 8

2	227	100.0	3311	11	AF039400	AF039400 Homo sapi
3	227	100.0	2826	42	AF127036	AF127036 Homo sapi
4	137	60.4	35278	11	AF039401	AF039401 Homo sapi
5	104.8	46.2	2937	12	AB017156	AB017156 Mus muscu
6	72.2	31.8	2984	3	BTU36445	BTU36445 Bos tauru
7	64.2	28.3	3317	3	AF001261	AF001261 Bos tauru
8	64.2	28.3	3288	3	AF001262	AF001262 Bos tauru
9	64.2	28.3	2820	3	AF001263	AF001263 Bos tauru
10	54.8	24.1	3022	12	AF047838	AF047838 Mus muscu
11	54.8	24.1	3471	12	AF052746	AF052746 Mus muscu
12	53.8	23.7	3415	42	AF043976	AF043976 Homo sapi
13	34	15.0	66736	9	HS262817	AL008709 Human DNA
14	34	15.0	99342	10	HS201017	AL020990 Human DNA
15	34	15.0	216991	34	HS164F3	AL035422 Homo sapi
16	33.8	14.9	130988	9	HS364122	AL031012 Human DNA
17	33.6	14.8	116210	11	HS406C18	AL023773 Human DNA
18	33.4	14.7	92800	11	AC004486	AC004486 Homo sapi
19	33.2	14.6	3604	9	AB026833	AB026833 Homo sapi
20	32.8	14.4	218956	34	CMS00008	AL049831 Homo sapi
21	32.2	14.2	62576	4	FRINEL1	AF146687 Fugu rube
22	32	14.1	191191	10	HSAC000119	AC000119 Human BAC
23	32	14.1	168644	11	AC004544	AC004544 Homo sapi
24	32	14.1	71353	11	AF104455	AF104455 Homo sapi
25	32	14.1	118610	34	AC002344	AC002344 Homo sapi
26	31.8	14.0	99014	11	HSU78027	U78027 Homo sapien
27	31.8	14.0	250529	11	HUAE000558	AE000558 Homo sapi
28	31.6	13.9	262929	35	AC007848	AC007848 Homo sapi
29	31.6	13.9	40594	36	CER03A10	269793 Caenorhabdi
30	31.6	13.9	24629	36	CER07D15	270685 Caenorhabdi
31	31.4	13.8	13405	7	SOYACCA	L48995 Glycine max
32	31.4	13.8	1825	36	AB065455	AB065455 Patinopec
33	31.2	13.7	118646	35	AC007173	AC007173 Drosophi
34	31.2	13.7	146502	42	AC005880	AC005880 Homo sapi
35	30.8	13.6	20297	7	SCCHXVILA	X94561 S.cerevisia
36	30.8	13.6	8989	7	SCFAS2	X76890 S.cerevisia
37	30.8	13.6	7164	7	SCFPL21W	X73587 S.cerevisia
38	30.8	13.6	7555	7	YSCFAS2A	J03936 S.cerevisia
39	30.8	13.6	245805	34	AC006752	AC006752 Caenorhab
40	30.8	13.6	200000	35	AC006229	AC006229 Homo sapi
41	30.8	13.6	38461	36	CEP38E11	268342 Caenorhabdi
42	30.4	13.4	9517	9	HUMINSP3R1	D26070 Human mRNA
43	30.4	13.4	1173	14	G05672	G05672 human STS W
44	30.4	13.4	137	14	G19634	G19634 human STS A
45	30.4	13.4	349	14	G30168	G30168 human STS S

## ALIGNMENTS

RESULT	1	
LOCUS	195746	878 bp DNA
DEFINITION	Sequence 8 from patent US 5733748.	
ACCESSION	195746	
NID	93940216	
VERSION	195746.1	GI:3940216

KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 878)
AUTHORS	Yu, G. and Rosen, C.
TITLE	Colon specific genes and proteins
JOURNAL	Patent: US 5733748-A 8 31-MAR-1998;
FEATURES	Location/Qualifiers
source	1..878

BASE COUNT	257 a	179 c	188 g	241 t	13 others
ORIGIN					
Query Match	100.0%	Score 227;	DB 5;	Length 878;	
Best Local Similarity	100.0%;	Pred. No. 1.9e-59;			

Matches	227	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	1	GTGGCGGCTCTGGAGAGATTAA	CGCAGCCAGACGAGAGATGAT	TATACCCACGACGAGATGA	60				
Db	56	GTGGCGGCTCTGGAGAGAGTTAA	CGCAGCCAGACGAGAGATGAT	TATACCCACGAGAGATGA	115				
QY	61	GCACGTGTCATACCGTGGGTGAT	TGAGATGATGATTAAGTAATCA	ATGAATGGAATCCACCAAGACCT	120				
Db	116	GCACGTGTCATACCGTGGGTGAT	TGAGATGATGATTAAGTAATCA	ATGAATGGAATCCACCAAGACCT	175				
QY	121	GAATTAATTAAGATGATGATGAT	TCAACACAGCAAGTGTGTTCA	GACAGAAATCTCTGGGA	180				
Db	176	GAATTAATTAAGATGATGATGAT	TCAACACAGCAAGTGTGTTCA	GACAGAAATCTCTGGGA	235				
QY	181	GGCTATTGTGGCTCTTGATGAT	CCCAAACTCTCCATACCTGATCT	227					
Db	236	GGCTATTGTGGCTCTTGATGAT	CCCAAACTCTCCATACCTGATCT	282					
RESULT	2	AF039400	3311 bp	mRNA	PrI	15-DEC-1998			
LOCUS		AF039400							
DEFINITION									
ACCESSION		AF039400							
NID		94009457							
VERSION		AF039400.1							
KEYWORDS									
SOURCE		human.							
ORGANISM		Homo sapiens							
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;							
AUTHORS		Eutheria; Primates; Catarrhini; Homnidae; Homo.							
TITLE		1 (bases 1 to 3311)							
		Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and							
		Pauli,B.U.							
		Genomic cloning, molecular characterization, and functional							
		analysis of human CACAI, the first human member of the family of							
		Ca2+-activated Cl- channel proteins							
		Genomics 54 (2), 200-214 (1998)							
		2 (bases 1 to 3311)							
		Gruber,A.D., Elble,R. and Pauli,B.U.							
		Direct Submission							
		Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,							
		Cornell University, Ithaca, NY 14853, USA							
FEATURES		Location/Qualifiers							
source		1..3311							
		/organism="Homo sapiens"							
		/db_xref="taxon:9606"							
		/chromosome="1"							
		/map="1p21-p31"							
		/tissue-type="small intestine"							
		1..3311							
		/gene="hCLCA1"							
		352..3096							
		/gene="hCLCA1"							
		/note="transmembrane glycoprotein"							
		/codon_start=1							
		/product="calcium-dependent chloride channel-1"							
		/protein_id="AAC93428.1"							
		/db_xref="PID:94009458"							
		/db_xref="GI:4009458"							
		/translation="MGPEKSVFILLHLLEGALSNLSLIQNNNGYDEGIYVAIDPNVP							
		EDPTLLOIKDWATQASIVLFPATGRPEYFKKVALIIPETWTKADYVRPKLETKYNA							
		DYLVATSTPRGDEPTPEBOMGNGCKGRERHILTPFIAGKLAEGPGKAFVEMAH							
		LRNGVDEYNDEKFTLSNGRIQAYRCSAGITGTNYVKKCGGSCYTKRCFNNVGL							
		YKEGCEFLVLOSRTKERSIMFAQHDVSIYVECTEONHKEAPKNOKNCRSTWETVL							

	BASE COUNT	1028 a	692 c	742 g	849 t	
ORIGIN						
Query Match						100.0%; Score 227; DB 11; Length 3311;
Best Local Similarity						100.0%; Pred. No. 1.9e-59;
Matches 227; Conservative						0; Mismatches 0; Indels 0; Gaps 0;
QY	1	GTGCGGGCTCTGGGAGGAGGTTAACCCAGCCAGCGAGAGTATATACCCAGCAGAGTGGG	60			
DB	2374	GTGCGGGGCTCTGGGAGGAGGTTAACCCAGCCAGCGAGAGTATATACCCAGCAGAGTGGG	2433			
QY	61	GCAGCTGACATACCGCGCGGCGATGAGATGATGAAATACATGGAATCCACCAAGACT	120			
DB	2434	GCAGCTGACATACCGCGCGGCGATGAGATGATGAAATACATGGAATCCACCAAGACT	2493			
QY	121	GAAATTAATAGAGATGATGTTCAACACAAGCAAGTGTGTTTACGACGAACAATCCTCGGGA	180			
DB	2494	GAAATTAATAGAGATGATGTTCAACACAAGCAAGTGTGTTTACGACGAACAATCCTCGGGA	2553			
QY	181	GGCTCATTTGGTGGCTTCGATGTCACCAATGTCCTCATCTGATCT	227			
DB	2554	GGCTCATTTGGTGGCTTCGATGTCACCAATGTCCTCATCTGATCT	2600			

RESULT	3
AF127036	
LOCUS	AF127036 2826 bp mRNA PRI 16-APR-1999
DEFINITION	Homo sapiens calcium-activated chloride channel protein 1 (CACCL)
ACCESSION	AF127036
NID	AF127036
KEYWORDS	mRNA, complete cds.
SOURCE	AF127036.1 GI:4585468
ORGANISM	. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 2826) Agnel, M., Vermaut, T. and Culouscou, J.-M. Cloning of three human homologs of bovine epithelial chloride channel Unpublished 2 (bases 1 to 2826) Agnel, M. and Culouscou, J.-M. Direct Submission Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des Carrieres, Rueil-Malmaison 92500, France Location/Qualifiers 1..2826 organism="Homo sapiens" db_xref="taxon:9606" feature_type="small intestine; colon" 1..2826 gene="CACCL" 5..2749 gene="CACCL1" codon_start=1 note="bovine epithelial chloride channel homolog" product="calcium-activated chloride channel protein 1" protein_id="AAD25487.1" db_xref="pid:g4585469" db_xref="gi:4585469" translation="MGPRKSSVFLLIHLLEGALNSLIOLNNNGEGIVVAIDPNVP EDETLIIQIKDAVTSALSLDEATGKRRIYFNVAAILPEMKTKADVRRLKEYNNA DLVASTPPGNDEPYLEOMGNCSEGRIRILTFDLAGKKLAIEYGSGCAFPVEMAH LMWGVEFDENDEKFYLSNIGRIQAFCASGITGVNVKCGGGSCITRKCTFKNVIGL

YKSGCEVLOSROTEKASIMFAOHVDSIVEFCTEONHNKEAPKONOKONLSTWEI  
RSEDFKRTTPTPTOPNPNTFSELOIGQATVCLVLDKSSMATGNRLNLANAGOLF  
LOTVELGSSVWGVTEFDSAAHVQSELIQINSQSDRDLARLPAAAGSGTISGLES  
FVIRKRYPTDSEIYLTDGNDNTISGCFNEVKOSGALIHVVALPNSAOBLELSK  
MTGGQATVSDOVNNGLIDAFGALSNGAVSORSIOELGVLALPSONMNGTIV  
DSTVGKDTLFTLWTTPOLILMDPSGKOGFVVDKTKMAVLOIPGIATVGTWY  
STOASSQTLTITVTSPASNAATLPITVTSKTKNDPSKPPSPLYVYANIROGASPLIRA  
SYTALIESYNGKTVTLELDNAGADATDDGVYSRYFTTYDINGRYSKVRALGCVN  
AARRVTPQOQSALYIPGMIENDEIOWNPPEINDVOHQVCFSTRSSGSEFVAS  
DVPNAPIDLPFPPOITDLKAIHGGSLINTWTAPGDYDHTGTAHXIIRISTSLD  
LKDKNESLQVNTALIPKEANSEVFLKPEPNTIFENGTDLFLAIQAVDKVLSKSEI  
SNIAVSLFIPQTPPETPSPDETSAPCPNHNINSTIPGIIHLKIMMWIGELQJLSIA

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN  
Query Match 100.0%; Score 227; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 1,9e-59;  
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCCGGGCTCTGGAGAGATTAAACGACGACGAGAGTGTATACCCAGCAGAGTGA 60  
|||||  
Db 2027 GTCCGGGCTCTGGAGAGATTAAACGACGACGAGAGTGTATACCCAGCAGAGTGA 2086  
QY 61 GCCTGTACTACTGCTGGCTGGATTGAGATGATGAATACATGGAATCCACCAACACT 120  
|||||  
Db 2087 GCCTGTACTACTGCTGGCTGGATTGAGATGATGAATACATGGAATCCACCAACACT 2146  
QY 121 GAATTAATATAGATGATGTTCAACACCAAGCATGTGTTTCAGCAGAACATCCTCGGGA 180  
|||||  
Db 2147 GAATTAATATAGATGATGTTCAACACCAAGCATGTGTTTCAGCAGAACATCCTCGGGA 2206  
QY 181 GGCTCATTTGTGGCTCTGTGATGTCCTCCCAATGCTCCCATACCTGATCT 227  
|||||  
Db 2207 GGCTCATTTGTGGCTCTGTGATGTCCTCCCAATGCTCCCATACCTGATCT 2253

RESULT 4  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS AF039401  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.  
ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1. 35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1. 1617  
/gene="hCLCA1"  
1. 33522  
gene

repeat\_region  
TATA\_signal  
exon  
mrna  
exon  
CDS  
/gene="hCLCA1"  
767..1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..24522,24628,26644..26859,27447..27708,28766,31840..32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPKRSVFLILHLLEGALNSLIQLNNNGEIVVAIDNPV  
EDRTILQOKMVTQASLXLPKATGRKRFKRVALLIPETWTKADYRPKLETYNAH  
DVLVASTPPEGDEFTPEOMGCGEKGERIHLTPFLAKRLAETGPOKARVEMAH  
LKGVEFDEVNDEKFIYLSNGRIQAVRCAGITGVNVAKKCGGSCYTRKFNKVTGL  
YKSGCEVLOSROTEKASIMFAOHVDSIVEFCTEONHNKEAPKONOKONLSTWEI  
RSEDFKRTTPTPTOPNPNTFSELOIGQATVCLVLDKSSMATGNRLNLANAGOLF  
LOTVELGSSVWGVTEFDSAAHVQSELIQINSQSDRDLARLPAAAGSGTISGLES  
FVIRKRYPTDSEIYLTDGNDNTISGCFNEVKOSGALIHVVALPNSAOBLELSK  
MTGGQATVSDOVNNGLIDAFGALSNGAVSORSIOELGVLALPSONMNGTIV  
DSTVGKDTLFTLWTTPOLILMDPSGKOGFVVDKTKMAVLOIPGIATVGTWY  
STOASSQTLTITVTSPASNAATLPITVTSKTKNDPSKPPSPLYVYANIROGASPLIRA  
SYTALIESYNGKTVTLELDNAGADATDDGVYSRYFTTYDINGRYSKVRALGCVN  
AARRVTPQOQSALYIPGMIENDEIOWNPPEINDVOHQVCFSTRSSGSEFVAS  
DVPNAPIDLPFPPOITDLKAIHGGSLINTWTAPGDYDHTGTAHXIIRISTSLD  
LKDKNESLQVNTALIPKEANSEVFLKPEPNTIFENGTDLFLAIQAVDKVLSKSEI  
SNIAVSLFIPQTPPETPSPDETSAPCPNHNINSTIPGIIHLKIMMWIGELQJLSIA

exon 27447..27708  
/gene="hclcal1"  
/number=12  
28766..28936  
/gene="hclcal1"  
/number=13  
31840..32079  
/gene="hclcal1"  
/number=14  
32919..33522  
/gene="hclcal1"  
/number=15

exon

BASE COUNT 11245 a 6870 c 7230 g 9933 t

ORIGIN

Query Match 60.4%; Score 137; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 6.5e-32;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 91 GATGAATACATGATGATCCACCAAGACCTGAATTAATAGATGATTCACACAG 150  
|||||  
Db 31839 GATGAATACATGATGATCCACCAAGACCTGAATTAATAGATGATTCACACAG 31838  
|||||

OY 151 CAAGTGTGTTACGACAGACATCCCTGGGAGGCTCATTTGTGCTGTGATGCCAAT 210  
|||||  
Db 31899 CAAGTGTGTTACGACAGACATCCCTGGGAGGCTCATTTGTGCTGTGATGCCAAT 31898  
|||||

OY 211 GCTCCATACCTGATCT 227  
|||||  
Db 31959 GCTCCATACCTGATCT 31958  
|||||

RESULT 5  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS AB017156 Mus musculus gob-5 mRNA, complete cds.  
ACCESSION AB017156  
NID 93721911  
VERSION AB017156.1 GI:3721911  
KEYWORDS GOB-5.  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (sites), Tanigawa, Y. and Hirohashi, S.  
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Cloning and identification of the gene gob-5, which is expressed in  
intestinal goblet cells in mice  
JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
MEDLINE 99160866  
REFERENCE 2 (bases 1 to 2937)  
AUTHORS Komiyama, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) to the DBJ/EMBL/GenBank databases. Tooru  
Komiyama, ERATO, JST, Hirohashi Cell Configuration Project; 5-9-9,  
Tokomai, Tsukuba, Ibaraki 300-2635, Japan  
(E-mail:tkom@hccp.jst.go.jp, Tel:81-298-47-7563,  
Fax:81-298-47-5226)  
FEATURES  
source location/Qualifiers  
1..2937  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15..2756  
/gene="gob-5"  
15..2736  
/gene="gob-5"  
CDS  
/gene="gob-5"  
15..2736  
/codon\_start=1  
/protein\_id="BAA3743.1"  
/db\_xref="pid:d1034712"

/db\_xref="pid:g3721912"  
/db\_xref="GI:3721912"  
/translation="MESLSPVFLILHLLEGLSLSLIQLNNNGIEGIVLIDBVP  
EDPALIOHKIDWTOASPYLFEATGRREFFKNVAILIPSMKAPETPTPKLETRNA  
DYLVSTSPGLGNDPEYTHEHAGCEGRIHLPLDPLACKLLOYGPODRTHBMAH  
FRMGVNEYNNDEKFPYSKGRPOAVGCSAITKNOVRNCGSGCTINQCVIYRVTG  
LYKNCVFPDPHONERKASIMENONINSVVEFTKNNHNOEAPNDONRCNLRSTVE  
IOESEDEKQTPMTPAOPPAFTSLUQIGQIVLVLDKSGSMINDRLRNMOASRLF  
LIQVTEQGSWGVNTPDSAAVVOSELKQNSGADRLILKHLPTVSGAGSTICGLRT  
AFYVIRKKYPTDGESEIVLIDGENDTISCFDLVKSGAIIHTVALGPAARETBSL  
KMTGGLQTSDDQVNNGLVDAPALSSGNAALQHSIOLESKVNLQNMONGSVI  
VDSVGKDTLELITWTHPTPIFIMDSGVQONGFILDITTKPVYASIPGTAVKGVK  
YSIASSQTLITVTSTRASATLPIVTEPVVKNKNGKFPSPVTVASIRQGSPLLR  
ASVATLISVVGKTVTELELDNAGADATKNDGYSHFFTAFPANGRYSVKIIALGV  
TSDRQRAAPFNKRAMTIDNTEDEDEYVMPNPREPYSVODKCFSTSSGGSFVATN  
VPAAPIPDLPFPCQITDLKASIOGQVNLVNTARPGDDYDHGRASVYIIRMSSTVYD  
LRDHFTSLQVNTGLIPKENSSEIEFEELGNTFENGDFITAIQAVDKSLKSEI  
SNIAVSVFIPAOEPPPEIDESTPCCPDISINSTIPGILHLIMKMLGEMQVTLGLH"

BASE COUNT 860 a 718 c 693 g 666 t

ORIGIN

Query Match 46.2%; Score 104.8; DB 12; Length 2937;  
Best Local Similarity 73.8%; Pred. No. 4.2e-22;  
Matches 149; Conservative 0; Mismatches 47; Indels 6; Gaps 1;

OY 5 GGGCTCGGAGAGGATTAACGACGAGAGAGTGAATACCCAGCAGAGTGGAGAC 64  
|||||  
Db 2044 GGGCTCGGAGAGGATTAACGACGAGAGAGTGAATACCCAGCAGAGTGGAGAC 2103  
|||||

OY 65 TGTACATACCTGGCTGATGAGATGATGAATTAACATGGAATCCACCAAGACCTGAAA 124  
|||||  
Db 2104 TGTACATACCTGGCTGATGAGATGATGAATTAACATGGAATCCACCAAGACCTGAAA 2163  
|||||

OY 125 TTAATAGATGATGATCAACACAGCAAGTGTGTTACAGACAGACATCCTCGGAGGCT 184  
|||||  
Db 2164 CTA-----GTTATGTCACAGACAGCAAGTGTGTTACAGACAGATCTTACGGGAGAT 2217  
|||||

OY 185 CATTGTGCTTGTGATGCC 206  
|||||  
Db 2218 CATTGTGCTTGTGATGCC 2239  
|||||

RESULT 6  
BTU36445 2984 bp mRNA NAM 09-FEB-1996  
LOCUS BTU36445 Bos taurus calcium-activated chloride channel mRNA, complete cds.  
ACCESSION U36445  
NID 91184065  
VERSION U36445.1 GI:1184065  
KEYWORDS COW.  
SOURCE Bos taurus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 2984)  
AUTHORS Cunningham, S.A., Awajda, M.S., Bubien, J.K., Ismailov, I.I.,  
Arrate, M.P., Berdiev, B.K., Benos, D.J. and Fuller, C.M.  
TITLE Cloning of an epithelial chloride channel from bovine trachea  
JOURNAL J. Biol. Chem. 270 (52), 31016-31026 (1995)  
MEDLINE 96125078  
REFERENCE 2 (bases 1 to 2984)  
AUTHORS Benos, D.J.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics,  
University of Alabama at Birmingham, BHSB 706, Birmingham, AL  
35294, USA  
FEATURES  
source location/Qualifiers  
1..2984  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="epithelial cells"





[illegible]

Db 2172 GATGTCGCAAGAAAGCCATGAGAACTGACATGCGAAGACCTTCACACAGGTAACTCTGGA 2231

QY 181 GGCTCATTTGGCTTCGATGTCGCAAAATGCTCCCATACCTGATC 226

Db 2232 GGGTCGTTACTGTGTCTGAGAGCGGCCCCCTATGGCGACACAGCTC 2277

RESULT 11

AF052746 AF052746 3471 bp mRNA ROD 08-APR-1999

LOCUS Mus musculus chloride channel Cacc mRNA, complete cds.

DEFINITION AF052746

ACCESSION AF052746

NID 93560546

VERSION AF052746.1 GI:3560546

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3471)

AUTHORS Romio,L., Musante,L., Cinti,R., Seri,M., Moran,O., Zegarra-Moran,O., and Galietta,L.J.V.

TITLE Characterization of a murine gene homologous to the bovine Cacc chloride channel

JOURNAL Gene 228 (1-2), 181-188 (1999)

MEDLINE 99173882

REFERENCE 2 (bases 1 to 3471)

AUTHORS Romio,L., Musante,L., Cinti,R., Seri,M., Moran,O., Seri,M. and Galietta,L.J.V.

TITLE Direct Submission

JOURNAL Submitted (09-MAR-1998) Lab. di Genetica Molecolare, Istituto Giannina Gaslini, Genova I-16148, Italy

FEATURES

source

1..3471

/organism="Mus musculus"

/db\_xref="taxon:10090"

/db\_xref="dbEST:AA245677"

47..2752

/codon\_start=1

/product="chloride channel Cacc"

/protein\_id="AAC35003.1"

/db\_xref="PIR:G3560547"

/db\_xref="GI:3560547"

/translation="MVGGLQVLLFLTLHLQNTSSMNVHLNSNGESGVIAINPSEDERLPIKEMVTOASTYLFPEASQGRVYFNISILVYPMTKSRPEYLPMPRESYKADVIYADPHLQGDDEPTLQVCGCGRGYIHETPELLTLNLRITGPRGFVHEMAHLRMGVETDYNDOFPYMSRKNTITERTSTITGNYVHNENRGNCYTRACRDSKTLIYERPCITIPKIQTAGASIMFMQNLNVSECTEKNNAEIPNLQNMCRNRSTWDYIKTSADFQNPAPMRGTEAPPPPTSLKSRRVCLVLDKSGSMKEDRLIRMQNAELYLIOYKESEMWGLVETDSAAHLQNYLIIKITSSTDYOKITANLPQOAGGTSICHGLOAGQATISDOSTSGSEIVLFDGEDNGIRSCFEAVRSQALIHITIALGPSAAELETLTSMQGLRLRYANKDNLSDILDFSRISTSGSVSOQLLESAFQVRAAGANGTVPALDSTVGDTFEVLITWYKREIILQDPKGRKTYTSDQDKLNIRSLQIPGTAEFGTWYSITGTSQILTMVTITRANSPTEPIPLATAHMSQSTRQYPSNIVYARVSQGLFTVIGANTALIEAHGHQVTLLEMDNAGADTVKNDQIVTRFYDGNRSYSLKRVDAQOQRNKRLLSLRQNKSLIYGVYENKIVLNPBPVOEALVATYEDENRVTSGGSEFTYSGAPDGDHARVPSKVTQLEAEFIDYHILMTAGKVLVDNGRAHRYIIRMSQHPDLIDEDENNATLVNASSLIPKEAGSEKFFKPEFFKFIANGIOLIALIQADNEASLTSEYSNIAQVNLKLTLEDSTLSALDISATLMTWGLTVITNSILN"

BASE COUNT 1096 a 777 c 736 g 862 t

ORIGIN

Query Match 24.1%; Score 54.8; DB 12; Length 3471;

Best Local Similarity 55.8%; Pred. No. 8e-07;

Matches 126; Conservative 0; Mismatches 97; Indels 3; Gaps 1;

QY 1 GTGGGGGCTGGGAGAGAGTTACGACGCCAGACGAGAGATACCCAGACAGAGTGA 60

Db 2084 GTGGCTGTCCGAGGCACAAAGAAACCAACGACATGTAGCTTAAGA---CAGAAGACAG 2140

QY 61 GCACGTACATACCTGGCTGAGATTGAGAATGATGAATATACAAATGAAATCCACCAAGACT 120

	Query Match	Similarity	24.18;	Score	54.8;	DB	12;	Length	3471;
	Best Local	Similarity	25.18;	Pred.	No.	8e-07;			
	Matches	126;	Conservative	0;	Mismatches	97;	Indels	3;	Gaps
									1;
QY	1	GTGGGGGCTCTGGGGAGGAGTTATACGCACCGCCAGAGGAGTGATATACCCACAGCAGTGA	60						
DB	2084	GTGGCTGTCTGGGGCAAGAAAGAAACCAAAACCCAGACTTGAAGCTTAAAG--CAAGAAACAG	2140						
QY	61	GCACGTATACATACCTGGCTGGTGATTAGAGATATGATGAATATACATGGAATCCACCAAGACT	120						

```

Db      2141 TCTTATATATACCTGCTATGTGAAATGTAATAATGTACTAGTAATCCACCCAGACCA 2200
Oy      121 GAATATATAGATGATGTCACACAGCAAGTGTGTTCAGACAGAACTCTCGGGA 180
Db      2201 GATGTCACAGAGAGAGCCATAGAACCTACAGTGGAAAGACTTCAACAGTAGTAACCTCGA 2260
Oy      181 GGCTCATTTGGCTTCGTGATGTCCCAATGCTCCCACTACCTGATC 226
Db      2261 GGGTCGTTACTGTGTCTGAGAGCCGCCCTGATGGCAGCACGCTC 2306

RESULT 12
AF043976 3415 bp mRNA PRI 07-APR-1999
LOCUS Homo sapiens CLCA homolog (hCLCA3) mRNA, complete cds.
DEFINITION AF043976
ACCESSION 94572288
NID AF043976.1 GI:4572288
VERSION AF043976.1 GI:4572288
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 3415)
AUTHORS Gruber A.D. and Pauli, B.U.
TITLE Molecular cloning and biochemical characterization of a truncated,
SECRETED member of the human family of Ca2+-activated Cl- channels
JOURNAL Biochim. Biophys. Acta 1444 (3), 418-423 (1999)
MEDLINE 99196715
REFERENCE 2 (bases 1 to 3415)
AUTHORS Gruber A.D., Elble, R.C. and Pauli, B.U.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1998) Department of Pathology, College of
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
FEATURES
source
1. 3415
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3415
/gene="hCLCA3"
15..807
/gene="hCLCA3"
/codon_start=1
/product="CLCA homolog"
/protein_id="AAD23734.1"
/db_xref="PIR:G4572289"
/db_xref="GI:4572289"
/translation="WFSLSKVIIEFLSLIPVKKSLVTLNNNGYDGIYIAINPSVE
DEKLIONIKEMVTEASTHLPHATKORAYFRANVSIILPMTYKSKSEYLLIPKQETYOAD
VIVADLYLKYGDDPYTLOYGCGDGKGYIHFPNELLNNLATYGRGKAVFPHGNAHL
RMGVFDEYNDQPFYISRNITETATRCSTRIYVYLVNLCAGASCIAARFRRDSOTGL
YEAKEFIPRSOTAKESIYFMONLDSVTEFCEYKTHNKEAPNL"
BASE COUNT 1149 a 625 c 630 g 1011 t
ORIGIN
Query Match 23.7%; Score 53.8; DB 42; Length 3415;
Best Local Similarity 56.5%; Pred. No. 1.6e-06;
Matches 100; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```

```

RESULT 13
HS262B17/c 66736 bp DNA PRI 25-NOV-1998
LOCUS Human DNA sequence from PAC HS262B17 on chromosome Xp11.23-Xp11.4.
DEFINITION Contains monomaline oxidase B (MAOB), ESTs, CA repeat.
ACCESSION AL008709
NID 92695811
VERSION AL008709.1 GI:2695811
KEYWORDS monomaline oxidase B; repeat polymorphism; Xp11.23-Xp11.4.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 66736)
AUTHORS Pearce, A.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1997) sanger.ac.uk/HGP/ChrX/ Sanger Centre,
Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Dec 18, 1997 this sequence version replaced gi:2598443.
IMPORTANT: This sequence is not the entire insert of clone 262B17.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre chromosome X
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 262B17 is at 1 in this sequence. The
true left end of clone 27K14 is at 66633.
262B17 is from the library RPII constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.
FEATURES
source
1. 66736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp11.23-Xp11.4"
/clone="262B17"
/clone_lib="RPII"
1..316
/note="WGA repeat: matches 90..426 of consensus"
813..1178
/note="THEIB repeat: matches 1..364 of consensus"
1575..1666
/note="LIMB8 repeat: matches 824..914 of consensus"
2052..2156
/note="LIMB2 repeat: matches 1055..950 of consensus"
complement(join(-2427..-2524,5934..5991,7594..7647,
8700..8796,20671..20830,22991..23140,24377..24518,
29424..29515,30552..30656,66119..>66256))
/codon_start=1
/product="MAOB"
/protein_id="CAA15489.1"
/db_xref="PIR:e1216621"
/db_xref="PIR:e1216621"
/db_xref="PIR:e1216621"
/db_xref="GI:2695812"
/translation="NOKKYVDLGSYVGPONRILRLAKELGLEYKVNEVERLIHH
VKGKSYPRGFPYVWNPITYLDHNNFRTDMDGREIPSDAPKAPLAEKMDMNTK

```





RESULT 14  
HS201D17 99342 bp DNA PRI 12-MAY-1999  
LOCUS Human DNA sequence from clone 201D17 on chromosome Xp11.4, complete  
DEFINITION  
ACCESSION  
AL020090  
93980348  
VERSION  
AL020090.1 GI:3980348  
KEYWORDS  
HTG  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 99342)  
TITLE  
McDougal, R.  
JOURNAL  
Direct Submission  
Submitted (12-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT  
On Dec 8, 1998 this sequence version replaced gi:2689203.  
requests: clone@sanger.ac.uk  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL.  
This sequence is the entire insert of clone 201D17. This sequence  
has been finished according to sequence map criteria as follows. An  
attempt is made to resolve all sequencing problems, such as  
compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome X, constructed by the Sanger Centre Chromosome X  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chrx  
201D17 is from the library RPII constructed at the Roswell Park  
Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/VECTOR: PCIPAC2.  
FEATURES  
Source  
1..99342  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="201D17"  
/clone\_id="RPII"  
/map="p11.4"  
1..225  
/note="L1MD3 repeat: matches 7515. .7738 of consensus"  
339..906  
/note="L1ME2 repeat: matches 5513. .6108 of consensus"  
938..1202  
/note="L1MA8 repeat: matches 6003. .6269 of consensus"  
1320..1651  
/note="MLT2FB repeat: matches 2. .360 of consensus"  
1667..1722  
/note="L14 copies 4 mer tctc 77% conserved"  
1667..1716  
/note="L25 copies 2 mer tc 80% conserved"  
1723..1816  
/note="L47 copies 2 mer ct 70% conserved"  
1730..1797  
/note="L17 copies 4 mer tcta 81% conserved"  
3020..3083  
/note="L2 repeat: matches 2679. .2750 of consensus"  
3398..3574  
/note="L1PA8 repeat: matches 5887. .6163 of consensus"

repeat\_region 5030..5108  
/note="L1PA2 repeat: matches 6066. .6144 of consensus"  
repeat\_region 5110..5185  
/note="L19 copies 4 mer agag 68% conserved"  
repeat\_region 5110..5179  
/note="L35 copies 2 mer ag 70% conserved"  
repeat\_region 7330..7463  
/note="L1R repeat: matches 103. .250 of consensus"  
repeat\_region 7936..8259  
/note="L2 repeat: matches 2433. .2749 of consensus"  
repeat\_region 8236..8738  
/note="L2 repeat: matches 1801. .2308 of consensus"  
repeat\_region 9509..9586  
/note="L2 repeat: matches 2635. .2710 of consensus"  
repeat\_region 9631..9696  
/note="L22 copies 3 mer ctc 79% conserved"  
repeat\_region 10533..11295  
/note="L2 repeat: matches 1. .790 of consensus"  
repeat\_region 11637..12073  
/note="L2 repeat: matches 2316. .2746 of consensus"  
repeat\_region 12076..12369  
/note="L2 repeat: matches 1. .295 of consensus"  
repeat\_region 12370..12616  
/note="L2 repeat: matches 2068. .2316 of consensus"  
repeat\_region 12911..13942  
/note="L1PA13 repeat: matches 5105. .6152 of consensus"  
repeat\_region 14352..14478  
/note="L1R3 repeat: matches 21. .124 of consensus"  
repeat\_region 14479..14757  
/note="L1R4 repeat: matches 16. .294 of consensus"  
repeat\_region 14758..14807  
/note="L1R5 repeat: matches 124. .178 of consensus"  
repeat\_region 15342..15406  
/note="L2 repeat: matches 2646. .2709 of consensus"  
repeat\_region 15507..15662  
/note="L2 repeat: matches 5. .158 of consensus"  
repeat\_region 15737..15783  
/note="L2 repeat: matches 2635. .2687 of consensus"  
repeat\_region 16120..16164  
/note="L2 repeat: matches 2644. .2687 of consensus"  
repeat\_region 16359..16487  
/note="L1MA9 repeat: matches 6142. .6271 of consensus"  
repeat\_region 16882..17332  
/note="L1R46 repeat: matches 1. .236 of consensus"  
repeat\_region 17586..17813  
/note="L1M4 repeat: matches 5604. .5820 of consensus"  
repeat\_region 17854..18287  
/note="L1R5 repeat: matches 1. .426 of consensus"  
repeat\_region 19173..19309  
/note="L1R1 repeat: matches 28. .170 of consensus"  
repeat\_region 20129..21086  
/note="L1R6 repeat: matches 1. .957 of consensus"  
repeat\_region 21087..21382  
/note="L1R7 repeat: matches 1. .296 of consensus"  
repeat\_region 21383..22840  
/note="L1R8 repeat: matches 957. .2418 of consensus"  
repeat\_region 23013..24262  
/note="L1P repeat: matches 2986. .4233 of consensus"  
repeat\_region 25925..26028  
/note="L26 copies 4 mer ctcc 65% conserved"  
repeat\_region 26579..26656  
/note="L2 repeat: matches 2623. .2700 of consensus"  
repeat\_region 26658..27100  
/note="L1 repeat: matches 4767. .5225 of consensus"  
repeat\_region 29019..29160  
/note="L1PA2 repeat: matches 5528. .5672 of consensus"  
repeat\_region 29161..29629  
/note="L1PA2 repeat: matches 5677. .6146 of consensus"  
repeat\_region 31344..31448  
/note="L1MA2 repeat: matches 6203. .6308 of consensus"  
repeat\_region 31611..31701  
/note="L1R repeat: matches 34. .129 of consensus"  
repeat\_region 31834..31925

```
repeat_region /note="L1MB8 repeat: matches 6077. .6167 of consensus"
33322. 32687
repeat_region /note="THE1B repeat: matches 1. .364 of consensus"
32718. 32775
repeat_region /note="L2 repeat: matches 2688. .2745 of consensus"
33184. 33590
repeat_region /note="MST4 repeat: matches 1. .426 of consensus"
34751. 35097
repeat_region /note="MER47A repeat: matches 20. .366 of consensus"
35479. 35670
repeat_region /note="MIR repeat: matches 19. .207 of consensus"
35677. 35783
repeat_region /note="MIR repeat: matches 3. .114 of consensus"
35784. 36230
repeat_region /note="L1R16A repeat: matches 1. .450 of consensus"
36231. 36284
repeat_region /note="MIR repeat: matches 114. .166 of consensus"
36389. 36506
repeat_region /note="MIR repeat: matches 25. .147 of consensus"
38244. 38595
repeat_region /note="L2 repeat: matches 2296. .2710 of consensus"
39242. 39342
repeat_region /note="MIR repeat: matches 49. .143 of consensus"
40293. 40679
repeat_region /note="MST4 repeat: matches 1. .394 of consensus"
40847. 41134
repeat_region /note="ALUSX repeat: matches 1. .304 of consensus"
41139. 41213
repeat_region /note="L2 repeat: matches 2465. .2534 of consensus"
41636. 41671
repeat_region /note="9 copies 4 mer at 86% conserved"
41637. 41672
repeat_region /note="18 copies 2 mer ta 89% conserved"
41673. 41947
repeat_region /note="L1P4S repeat: matches 5871. .6142 of consensus"
42955. 43008
repeat_region /note="MERSB repeat: matches 4. .58 of consensus"
43285. 43573
repeat_region /note="ALUSX repeat: matches 1. .289 of consensus"
43698. 43879
repeat_region /note="L1R16C repeat: matches 199. .387 of consensus"
44140. 44175
repeat_region /note="9 copies 4 mer tta 83% conserved"
47020. 47165
repeat_region /note="MIR repeat: matches 3. .156 of consensus"
47371. 47987
repeat_region /note="L2 repeat: matches 2129. .2742 of consensus"
48002. 48366
repeat_region /note="L1R16A repeat: matches 17. .445 of consensus"
48382. 48502
repeat_region /note="L2 repeat: matches 2004. .2121 of consensus"
48521. 48673
repeat_region /note="L2 repeat: matches 1609. .1759 of consensus"
49754. 49887
repeat_region /note="L2 repeat: matches 2359. .2483 of consensus"
50900. 50949
repeat_region /note="25 copies 2 mer tt 72% conserved"
51672. 51724

Query Match 15.0%; Score 34; DB 10; Length 99342;
Best Local Similarity 50.6%; Pred. No. 2;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
```

```
Db 22924 AAGAGCTTTGAGAGTCATGTTGAATAGAGGTTGGAGCT 22965
RESULT 15
HS164F3
LOCUS HS164F3 216991 bp DNA HTG 11-JUN-1999
DEFINITION Homo sapiens chromosome X clone 164F3, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION AL035422
NID 95051831
VERSION AL035422.11 GI:5051831
KEYWORDS HTG: HTCS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 216991)
REFERENCE Wilson, S.
AUTHORS Direct Submission
TITLE Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 12, 1999 this sequence version replaced gi:5050941.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: d164F3 Contig_ID: 03649 acc=AL035422
length: 216991 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source 1. 216991
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="164F3"

BASE COUNT 63201 a 44436 c 46495 g 62843 t 16 others
ORIGIN
```

```
Query Match 15.0%; Score 34; DB 34; Length 216991;
Best Local Similarity 66.2%; Pred. No. 2;
Matches 49; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 83 TTGAGATGATGAAATACATGATCCACCAAGACCTGAATTAATAGATGATGTC 142
DB 125283 TTGATATCATATAACTTAAGTCAATACAGCTAGACATGAAGGAAATATG 125342
QY 143 AACCAAGCAAGTG 156
DB 125343 AACCCAAAAAATG 125356
```

Search completed: August 5, 1999, 23:14:23  
Job time: 7746 sec

---

This Page Blank (uspto)



KM diagnostic assay: colon cancer; antibody; screening; ss.  
OS Homo sapiens. Location/Qualifiers  
FH Key 2.685  
FT CDS /tag= a  
FT /note= "no stop codon given"  
PN U5573748-A.  
PD 31-MAR-1998.  
PE 06-JUN-1995: 469667.  
PR 06-JUN-1995: US-469667.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen C, Yu G;  
WP: 98-229823/20.  
DR P-PSDB: M46879.  
PT Colon-specific nucleic acids - useful as probes for detecting colon  
cancer micrometastases  
PS Claim 15: Fig 5A-B; 51pp; English.  
CC V16668-81 represent polynucleotide sequences of partial or full-length  
cDNA clones of colon-specific genes. The polynucleotides can be used  
as probes to detect expression of the corresponding human genes, e.g. in  
diagnostic assays containing the polynucleotides of colon cancer.  
CC Recombinant cells containing the polynucleotides can be used to  
produce the polypeptides, in order that antibodies can be raised and  
used in further screening or diagnostics.  
SQ Sequence 878 BP; 257 A; 179 C; 188 G; 241 T;

Query Match 100.0%; Score 227; DB 1; Length 878;

Best Local Similarity 100.0%; Pred. No. 5.7e-67;  
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTGCGGCTCTGGAGAGTTACGACGACGAGAGATGATACCCGACGAGATGGA 60  
DB 56 GTGCGGCTCTGGAGAGTTACGACGACGAGAGATGATACCCGACGAGATGGA 115  
OY 61 GCATCTGATACCTCTGGGATTTGAGATGATGAATTCATGGAATCCACGAGACT 120  
DB 116 GCATCTGATACCTCTGGGATTTGAGATGATGAATTCATGGAATCCACGAGACT 175  
OY 121 GAAATTAATAGGATGATGTCACACAGCAAGTGTGTTGAGAGAAATCCTCGGGA 180  
DB 176 GAAATTAATAGGATGATGTCACACAGCAAGTGTGTTGAGAGAAATCCTCGGGA 235  
OY 181 GGCTCATTTGTGGCTTGTGATGTCCCAATGCTCCATACCTGATCT 227  
DB 236 GGCTCATTTGTGGCTTGTGATGTCCCAATGCTCCATACCTGATCT 282

## RESULT 3

T58840\_5/c  
Continuation (6 of 6) of T58840 from base 500001 (Mycoplasma genitalium genome.)  
WP Sequence split into 6 fragments LOCUS T58840 Accession T58840

WP	Fragment Name	Begin	End
WP	T58840_0	1	110000
WP	T58840_1	100001	210000
WP	T58840_2	200001	310000
WP	T58840_3	300001	410000
WP	T58840_4	400001	510000
WP	T58840_5	500001	580073

Query Match 13.0%; Score 29.4; DB 1; Length 80073;  
Best Local Similarity 52.9%; Pred. No. 10;  
Matches 63; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY 63 ACCTGATACCTGGGCTGATGAGATGATGAATACATGGAATCCACGAGACTGA 122  
DB 79458 AGTTAAGCTTCTGACTGCTTATATATGATGATGATTTGCTAAACAAACTAA 79399  
OY 123 AATTAATAGGATGATGTTCAACACAGCAAGTGTGTTGAGAGAAATCCTCGGAG 181  
DB 79398 AAAAAATATGATGCTTTTCTAAATAATTTGTTCTGACATCAAAATCATCAGCAG 79340

RESULT 4  
X37374/c  
ID X37374 standard; cDNA; 1146 BP.

AC X37374.  
DT 06-JUL-1999 (first entry)  
DE Human secreted protein cDNA fragment containing gene 6.  
KW Human; secreted protein; prevention; treatment; protein therapy;  
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;  
KW developmental abnormality; foetal deficiency; blood disorder; lymphoma;  
KW leukemia; immune system disorder; autoimmune disease; hepatic disease;  
KW renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;  
KW Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;  
KW arthritis; psoriasis; digestive; endocrine; infection; ss.

OS Homo sapiens.  
PN M09090155-A1.  
PD 25-FEB-1999.  
PE 18-AUG-1998: U17044.  
PR 15-JUL-1998: US-092956.  
PR 15-JUL-1998: US-092956.  
PR 15-AUG-1997: US-056368.  
PR 19-AUG-1997: US-056369.  
PR 19-AUG-1997: US-056535.  
PR 19-AUG-1997: US-056535.  
PR 19-AUG-1997: US-056556.  
PR 19-AUG-1997: US-056628.  
PR 19-AUG-1997: US-056629.  
PR 19-AUG-1997: US-056726.  
PR 19-AUG-1997: US-056726.  
PR 19-AUG-1997: US-056728.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Brewer LA, Duan R, Ebner R, Endress GA, Feng P,  
PI Florence C, Florence KA, Komatsoulis GA, Lafleur DM,  
PI Moore PA, Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR,  
PI Young PE;  
DR WPI: 99-190160/16.  
DR P-PSDB: Y07749.

PT New isolated human genes and the secreted polypeptides they encode  
PT - useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
PS Claim 1a: Page 193-194, 280pp; English.  
CC This invention describes novel isolated human secreted proteins and  
CC their encoding nucleic acid sequences. The products of the invention  
CC are useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. Also pathological conditions can be  
CC diagnosed by determining the presence or amount of expression of  
CC the new polypeptides in a sample or by determining the presence or  
CC absence of mutations in the new polynucleotides. Specific uses are  
CC described for each of the 70 polynucleotides, based on which tissues they  
CC are most highly expressed in, and include developing products for the  
CC diagnosis or treatment of cancer, tumours, neurodegenerative  
CC disorders, developmental abnormalities and foetal deficiencies, blood  
CC disorders, leukemias, diseases of the immune system, autoimmune diseases,  
CC hepatic and renal disease, lymphomas, inflammation, allergic asthma,  
CC sepsis, diabetes, Alzheimer's and cognitive disorders, schizophrenia,  
CC osteoporosis, arthritis, psoriasis, digestive/endocrine disorders,  
CC infections and AIDS. The human secreted proteins of the invention are  
CC represented in Y07744-Y07850 and the encoding nucleic acids are  
CC represented in X37369-X37441.  
SQ Sequence 1146 BP; 349 A; 171 C; 177 G; 447 T;

Query Match 13.0%; Score 29.4; DB 1; Length 1146;  
Best Local Similarity 58.6%; Pred. No. 1.6;  
Matches 51; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 68 ACATACCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 127  
DB 724 AGATCCCTGATGAGGATTTAAGATGATGATGATGATGATGATGATGATGATGAT 675  
OY 128 AATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 154  
DB 674 ACCCAATCATTTTCAACACAGCAAG 648

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 23:10:15 ; Search time 3264.21 Seconds

(without alignments)  
176.348 Million cell updates/sec

Title: US-09-049-696-4

Perfect score: 181

Sequence: 1 CAAAAGATGCACATTCATTA.....ACAAGCAAAATCAAAATATCC 181

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBank: \*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_v1:\*  
18: em\_fun:\*  
19: em\_hcg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_v1:\*  
34: gb\_hcg1:\*  
35: gb\_hcg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gp\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	179.4	99.1	3311	11	AF039400	Homo sapi

2	179.4	99.1	2826	42	AF127036	AF127036 Homo sapi
3	115.4	63.8	35278	11	AF039401	AF039401 Homo sapi
4	94.8	52.4	2937	12	AB017156	AB017156 Bos taurus
5	74.2	41.0	3317	3	AF001261	AF001261 Bos taurus
6	74.2	41.0	3288	3	AF001262	AF001262 Bos taurus
7	74.2	41.0	2820	3	AF001263	AF001263 Bos taurus
8	74.2	41.0	1265	3	AF001264	AF001264 Bos taurus
9	72.8	40.2	3415	42	AF043976	AF043976 Homo sapi
10	71.2	39.3	3022	12	AF047838	AF047838 Mus muscu
11	67.8	37.5	2984	4	AF052746	AF052746 Mus muscu
12	58.8	32.5	3604	9	AB026833	AB026833 Homo sapi
13	58.8	32.5	3604	9	AB026833	AB026833 Homo sapi
14	38.4	21.2	1691	4	CHK019A2M	J00838 Chicken typ
15	38.4	21.2	1648	4	CHK00LA29	M25984 Chicken alp
16	38.4	21.2	1683	4	GCC00LA29	V00390 Messenger f
17	38.4	21.2	1010	4	GCC00LA29	X01614 Chicken pro
18	37	20.4	251199	11	H0A0000659	AE000659 Homo sapi
19	36.8	20.3	94695	7	ATF23E13	AT022141 Arabidops
20	36.8	20.3	8700	8	AF068574	AF068574 Arabidops
21	36.2	20.0	9432	37	AE001375	AE001375 Plasmodiu
22	35.2	19.4	1056	12	M050UNKB	L04849 Mouse (clon
23	35.2	19.4	113899	36	PFMAJ3P4	AL008970 Plasmodiu
24	34.8	19.2	1691	8	AF028338	AF028338 Arabidops
25	34.6	19.1	214550	34	CNS000009	AL049830 Homo sapi
26	34.4	19.0	105127	11	AC004548	AC004548 Homo sapi
27	34.2	18.9	160282	35	AC006996	AC006996 Homo sapi
28	34.2	18.9	2379	36	DD31	X54452 D.discoiden
29	33.8	18.7	21618	9	HS65B7	Z83849 Human DNA s
30	33.8	18.7	180104	11	DJ526N18	AF123462 Homo sapi
31	33.8	18.7	142156	34	HSDJ685D6	AL049642 Homo sapi
32	33.6	18.6	110000	34	CEY105E8_0	AL022594 Caenorhab
33	33.4	18.5	153835	34	CEY6D1	AL023673 Caenorhab
34	33.4	18.5	28999	36	CEY44E5	Z83108 Caenorhabd1
35	33.2	18.3	110514	8	F7G19	AC000106 Sequence
36	33.2	18.3	2312	36	TCG93689	U93689 Toxoplasma
37	33.2	18.3	29149	37	CELK09E2	U23527 Caenorhabd1
38	33.2	18.2	132783	11	AC000378	AC000378 Human chr
39	33	18.2	5513	12	D65733	D65733 Mus muscu
40	33	18.2	158405	12	NMHC310M6	AF109906 Mus muscu
41	33	18.2	110000	34	CEY38E10_0	AL021149 Caenorhab
42	33	18.2	296699	34	CEY48A6	Z92854 Caenorhabd1
43	33	18.2	75073	36	CEY48A6B	AL023844 Caenorhab
44	32.8	18.1	777	10	HSFKBMR	X52220 Human FKBP
45	32.8	18.1	29031	36	CEM05B10	Z74476 Caenorhabd1

## ALIGNMENTS

RESULT	1	AF039400	3311 bp	mrna	PRI	15-DEC-1998
LOCUS		AF039400				
DEFINITION		Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA, complete cds.				
ACCESSION		AF039400				
NID		94009457				
VERSION		AF039400.1	GI:4009457			
KEYWORDS						
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		1 (bases 1 to 3311)				
AUTHORS		Gruber,A.D., Elbie,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.				
TITLE		Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins				
JOURNAL		Genomics 54 (2), 200-214 (1998)				
MEDLINE		99047526				
REFERENCE		2 (bases 1 to 3311)				
AUTHORS		Gruber,A.D., Elbie,R. and Pauli,B.U.				
TITLE		Direct Submission				

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES  
source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue-type="small intestine"  
1. .3311  
/gene="hCLCA1"  
352. .3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="PID:94009458"  
/db\_xref="GI:4009458"  
/translation="MGPKSSVFILILHLLEGALSNSLIQINNNNGEIVVAIDPNVP  
EDFTLQIKMVTQASLYLEFATGKRFYKFNVAIILPEWTKADYRPRLETYNNA  
DYVAESTPRENDPEYTEMONGEGERHILPDLTAGKRLAEYGPQKAFVEMAH  
LRMGVEYNDDEKFLYLSNGRIQAVCSAGITGTIVYKCCGSCCYTRCFENVTGL  
YKGECEFLVQSRQTEKASIMFRAQVDSIVECTEONHNKAPKONKCNLRSTWEVI  
RDSDEKFTTPTTQPPNPTSLQIGRIYCLVLDKSGSMATGNRLNRLNQGLFL  
LQYELSGWGVTFDSAAHVOSELQINSQSDRDLAKRLPAAASGGSISCGLSRA  
FTVIRKRYPTDSEIVLTDGEDNTISCFENKQSGAIHTVALGPSAAOLELELSK  
MTGICLOTASDOVONNGILDAFGLSSGNQAVSQRISLESKGLTIQNSOMNGTVIY  
DSYVKRDLFLITMTTQPPQILMDPSGKGGFVDKNTKAVLIQIGIAKGVTKY  
SLQASSQTLTLTVTSRASNAITLPIYVTSKTKNDTSKPSPLVYANRQASPLIRA  
SVTALIESVNGKTVLELLDNGADATKDDVSRITTTIDNKRISYKRALGAVN  
AARRVYIPQOSGALYIPGNIENDEIQMNPPEIKNDVQKRVCFSTSSGGSVAS  
DVNPAPIDLPFGQITDLKAEIHGSLINLMTAPDGDYDHTAHKYLIRISITLD  
LRDKFNSLIQVNTTALIPKANSSEVLEFPENITFENGTDLFLIAIQADVLDKSEI  
SNIRAVSLFIPQTPPTPSPDETSAFCPNHINSTITPGIHLIKIMMKWIGELQSLA  
"

BASE COUNT 1028 a 692 c 742 g 849 t

ORIGIN

Query Match 99.1%; Score 179.4; DB 11; Length 3311;  
Best Local Similarity 99.4%; Pred. No. 2.8e-34;  
Matches 180; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAAGATGCACATTAAGTAACAGACCTATGAAAAAGATGTGAGTTGTCT 60  
DB 972 CAAAAGATGCACATTAAGTAACAGACCTATGAAAAAGATGTGAGTTGTCT 1031  
QY 61 CCAATCCCGCCAGACGAGAGGCTTCTATATGTTTGCACAAACATGTTCTATAGT 120  
DB 1032 CCAATCCCGCCAGACGAGAGGCTTCTATATGTTTGCACAAACATGTTCTATAGT 1091  
QY 121 TGAATTCGTACAGAACCAACCAACCAAGAGCTCCAAACAGCAAAATCAAAATG 180  
DB 1092 TGAATTCGTACAGAACCAACCAACCAAGAGCTCCAAACAGCAAAATCAAAATG 1151  
QY 181 C 181  
DB 1152 C 1152

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL1)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue-type="small intestine; colon"  
1. .2826  
/gene="CACCL1"  
5. .2749  
/gene="CACCL1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPKSSVFILILHLLEGALSNSLIQINNNNGEIVVAIDPNVP  
EDFTLQIKMVTQASLYLEFATGKRFYKFNVAIILPEWTKADYRPRLETYNNA  
DYVAESTPRENDPEYTEMONGEGERHILPDLTAGKRLAEYGPQKAFVEMAH  
LRMGVEYNDDEKFLYLSNGRIQAVCSAGITGTIVYKCCGSCCYTRCFENVTGL  
YKGECEFLVQSRQTEKASIMFRAQVDSIVECTEONHNKAPKONKCNLRSTWEVI  
RDSDEKFTTPTTQPPNPTSLQIGRIYCLVLDKSGSMATGNRLNRLNQGLFL  
LQYELSGWGVTFDSAAHVOSELQINSQSDRDLAKRLPAAASGGSISCGLSRA  
FTVIRKRYPTDSEIVLTDGEDNTISCFENKQSGAIHTVALGPSAAOLELELSK  
MTGICLOTASDOVONNGILDAFGLSSGNQAVSQRISLESKGLTIQNSOMNGTVIY  
DSYVKRDLFLITMTTQPPQILMDPSGKGGFVDKNTKAVLIQIGIAKGVTKY  
SLQASSQTLTLTVTSRASNAITLPIYVTSKTKNDTSKPSPLVYANRQASPLIRA  
SVTALIESVNGKTVLELLDNGADATKDDVSRITTTIDNKRISYKRALGAVN  
AARRVYIPQOSGALYIPGNIENDEIQMNPPEIKNDVQKRVCFSTSSGGSVAS  
DVNPAPIDLPFGQITDLKAEIHGSLINLMTAPDGDYDHTAHKYLIRISITLD  
LRDKFNSLIQVNTTALIPKANSSEVLEFPENITFENGTDLFLIAIQADVLDKSEI  
SNIRAVSLFIPQTPPTPSPDETSAFCPNHINSTITPGIHLIKIMMKWIGELQSLA  
"

BASE COUNT 875 a 623 c 632 g 696 t

ORIGIN

Query Match 99.1%; Score 179.4; DB 42; Length 2826;  
Best Local Similarity 99.4%; Pred. No. 2.9e-34;  
Matches 180; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAAGATGCACATTAAGTAACAGACCTATGAAAAAGATGTGAGTTGTCT 60  
DB 625 CAAAAGATGCACATTAAGTAACAGACCTATGAAAAAGATGTGAGTTGTCT 684  
QY 61 CCAATCCCGCCAGACGAGAGGCTTCTATATGTTTGCACAAACATGTTCTATAGT 120  
DB 685 CCAATCCCGCCAGACGAGAGGCTTCTATATGTTTGCACAAACATGTTCTATAGT 744  
QY 121 TGAATTCGTACAGAACCAACCAACCAAGAGCTCCAAACAGCAAAATCAAAATG 180  
DB 745 TGAATTCGTACAGAACCAACCAACCAAGAGCTCCAAACAGCAAAATCAAAATG 804  
QY 181 C 181  
DB 805 C 805

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
DEFINITION complete cds.

Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 2826)  
Agnel, M., Vernalt, T. and Culouscou, J.-M.  
cloning of three human homologs of bovine epithelial chloride  
channel  
Unpublished  
2 (bases 1 to 2826)  
Agnel, M. and Culouscou, J.-M.  
Direct Submission  
Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Rueil-Malmaison 92500, France  
Location/Qualifiers  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue-type="small intestine; colon"  
1. .2826  
/gene="CACCL1"  
5. .2749  
/gene="CACCL1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPKSSVFILILHLLEGALSNSLIQINNNNGEIVVAIDPNVP  
EDFTLQIKMVTQASLYLEFATGKRFYKFNVAIILPEWTKADYRPRLETYNNA  
DYVAESTPRENDPEYTEMONGEGERHILPDLTAGKRLAEYGPQKAFVEMAH  
LRMGVEYNDDEKFLYLSNGRIQAVCSAGITGTIVYKCCGSCCYTRCFENVTGL  
YKGECEFLVQSRQTEKASIMFRAQVDSIVECTEONHNKAPKONKCNLRSTWEVI  
RDSDEKFTTPTTQPPNPTSLQIGRIYCLVLDKSGSMATGNRLNRLNQGLFL  
LQYELSGWGVTFDSAAHVOSELQINSQSDRDLAKRLPAAASGGSISCGLSRA  
FTVIRKRYPTDSEIVLTDGEDNTISCFENKQSGAIHTVALGPSAAOLELELSK  
MTGICLOTASDOVONNGILDAFGLSSGNQAVSQRISLESKGLTIQNSOMNGTVIY  
DSYVKRDLFLITMTTQPPQILMDPSGKGGFVDKNTKAVLIQIGIAKGVTKY  
SLQASSQTLTLTVTSRASNAITLPIYVTSKTKNDTSKPSPLVYANRQASPLIRA  
SVTALIESVNGKTVLELLDNGADATKDDVSRITTTIDNKRISYKRALGAVN  
AARRVYIPQOSGALYIPGNIENDEIQMNPPEIKNDVQKRVCFSTSSGGSVAS  
DVNPAPIDLPFGQITDLKAEIHGSLINLMTAPDGDYDHTAHKYLIRISITLD  
LRDKFNSLIQVNTTALIPKANSSEVLEFPENITFENGTDLFLIAIQADVLDKSEI  
SNIRAVSLFIPQTPPTPSPDETSAFCPNHINSTITPGIHLIKIMMKWIGELQSLA  
"



ACCESSION AF039401  
NTD 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human hCICAI, the first human member of the family of Ca<sup>2+</sup>-activated Cl<sup>-</sup> channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source 1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCICAI"  
1..33522  
/gene="hCICAI"  
767..1265  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587..1596  
/gene="hCICAI"  
/note="putative"  
1618..1708  
/gene="hCICAI"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,22522..24628,26644..26859,27447..27708,28766..28936,28766..28936,31840..32079,32919..33522)  
/gene="hCICAI"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCICAI"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)  
/gene="hCICAI"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPEKSSVFILILHLEGALSNSLIQIANNNGYEGIVAVDPNV  
EDFIIIOIKMVAOASLIFEAATGKRFYENVALILPETKTRADYVRPLRYEMAH  
DLVAESTPRNDPEYTEOMNGCEGKRRIHLTDFIAGKLAERYGPKRFVEMAH  
LRMGVDEYNDKFIYLSNGRIQAVRCAGTGNVAKCGGSGCTTKRCFENVTGL  
YEKGEFVLQROTETKASIMFAQVDSIVECTQNNKRNKPNQONCINRSTWEVI  
RDSDFKTKTPMTQPPNPFTSLQIGRIYCLVATGSGSNATGRLNRLNQQLFL  
LQYELGSGVAMTFEDSAAYOSELIQINSQSDYRTAKRLPAASAGRTSICGLRSLA  
FTVIRKRPYDGESEVILITDGEINTEISCFENVSQGLITHTVALGPEAAQELRSLK  
MTGSLQTYASQVONNGCIDAFGLSSNGAVSQRISOLESGKGLIOWSONMNGTVY  
DSIVGKDTLFLITWTPQILMPDPSQKQKGFVNDNTMALIQIGIKAVGWIKY  
SLQASQDTLLITVSRASNATLPITVTSKNGKDKSRPSPLVYVANTROASPLIRA  
SVTALIESVNGKTYVLELDNGAGADATKDGVSRYRTYTDINGRYSVKRALGVN  
AARRVIEPOOGALYIPGMINDEIQNPPPEIKDVQHKQVCFSTSSGSGFEVAS  
DVNAPIPDLPFGQITDNLNIEHGSLINLTWAPAGDDYHGHAHKYIIINISIID

LRDKENSIOVNTALIPKEANSEEVLPKRENTFENGTDLFIAIOAVDKLDSKEI  
SNIRVSLFIPQTPPETPSPDEISACPNIHINSTIPGHIILKIMKNGIGELDLSIA  
6667..6807  
/gene="hCICAI"  
/number=3  
6979..7126  
/gene="hCICAI"  
/number=4  
9696..9801  
/gene="hCICAI"  
/number=5  
15454..15631  
/gene="hCICAI"  
/number=6  
18591..18809  
/gene="hCICAI"  
/number=7  
19783..20010  
/gene="hCICAI"  
/number=8  
22253..22427  
/gene="hCICAI"  
/number=9  
24522..24628  
/gene="hCICAI"  
/number=10  
26644..26859  
/gene="hCICAI"  
/number=11  
27447..27708  
/gene="hCICAI"  
/number=12  
28766..28936  
/gene="hCICAI"  
/number=13  
31840..32079  
/gene="hCICAI"  
/number=14  
32919..33522  
/gene="hCICAI"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 63.8%; Score 115.4; DB 11; Length 35278;  
Best Local Similarity 91.7%; Pred. No. 4.7e-19;  
Matches 122; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 1 CAAAAGATGCACATTCATTAAGTACAGACACTATGAAAAAGATGTGAGTTGTTCT 60  
DB 15517 CAAAAGATGCACATTCATTAAGTACAGACACTATGAAAAAGATGTGAGTTGTTCT 15576  
QY 61 CCAATCCCGCAGAGGAGGCGCTTCATATGTTGCACAACATGTTGATTCATAGT 120  
DB 15577 CCAATCCCGCAGAGGAGGCGCTTCATATGTTGCACAACATGTTGATTCATAGT 15566  
QY 121 TGAATTCGTACA 133  
DB 15637 TACCTGTCTCA 15649  
RESULT 4  
AB017156 2937 bp mRNA ROD 04-MAR-1999  
LOCUS AB017156  
DEFINITION Mus musculus gob-5 mRNA, complete cds.  
ACCESSION AB017156  
VERSION 93721911  
KEYWORDS AB017156.1 GI:3721911  
SOURCE GDB-5.  
ORGANISM Mus musculus adult intestine goblet cell cDNA to mRNA.

REFERENCE 1 (sites) Komiya, T., Tanigawa, Y. and Hirohashi, S.  
AUTHORS Cloning and identification of the gene gob-5, which is expressed in  
TITLE intestinal goblet cells in mice  
JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
MEDLINE 99160866  
REFERENCE 2 (bases 1 to 2937)  
AUTHORS Komiya, T., Tanigawa, Y. and Hirohashi, S.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/Genbank databases. Tchrn  
Komiya, ERATO, JST, Hirohashi Cell Configuration Project; 5-9-9,  
Tokodai, Tsukuba, Ibaraki 300-2635, Japan  
(E-mail: tkomehccp.jst.go.jp, Tel:81-298-47-7563,  
Fax:81-298-47-5226)

FEATURES  
source  
1. .2937  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
15. .2756  
/gene="gob-5"  
15. .2756  
/gene="gob-5"  
/codon\_start=1  
/protein\_id="AAA33743.1"  
/db\_xref="PID:d1034712"  
/db\_xref="PID:g3721912"  
/db\_xref="GI:3721912"  
/translation="MESLSPVFLILHLLEGVLESILQINNGEGIVAIIDHVP  
EDELIOHKMTQASPLPEATGKRFYKNAVILIPESWAKPEYTRLETFKNA  
DVLVSTRPLNDPRTYEHIGACGCKIRILTPDLAKGLIYODPRTYEHMAH  
FRWCVFNEDNDEKFTYSKRPQAVRCASALTGKNVRRCGSGCINGCVDIVRNG  
LYKNCVFPDPHONERKASIMENONINSVETCEKNNDAPNDONRCALSTWY  
IOESDEFKOTPMTAOPAPYSLQIGQIVCLDLKSGSMLMDRLNRNQAISRL  
LLOVEGSGVMTFDSAAYOSLQKNSGADRLIKHLPVSAGSGSICGLT  
AFYIKKKYPTDGESEIYLDGEDMTISCDLVKSGALIHITVALGPAAKAELOS  
KMTGICQVSSDOYONGNIDAPFALSNGAIAOHSTOLESRGVNLONNOMNGSVY  
VDSVGRDTELLITWTRPTITFTDPSGVONKFIIDTTKVAIVLOVPGAKYFVK  
YSIASSTQLTLVTSRAKATLEPTITVFNKNTGCFPSYVYASIRGASPIIR  
ASVALIESVNGKVTLELLNGAGADATKNDGVSRFLAFDANGRSVAKMLGCV  
TSDRRAAPKRNAMIDGWIJEDGVRNMPRPETSYQDKQLCFSTRISGGSVAVN  
VPAAPAIIDLPPOITDLKASIGONLVNTLWAPGDYDHGRASNTIIMSTSIYD  
LRDHNTSLQVNTGLIPKEASSEIEFEIGNTFGNGDTIFALQAVDSNLSKEI  
SNIAVSVFITRAQPRIPEDSTPPCDISINSITIPGIVHLIMKMLGEMQVTLGLH"

BASE COUNT 860 a 718 c 693 g 666 t

ORIGIN

Query Match 52.4%: Score 94.8; DB 12; Length 2937;  
Best Local Similarity 70.8%; Pred. No. 6.3e-14;  
Matches 126; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

4 AAGATGCACATTCATTAAGTATACAGAGACTCTATGAAAAAGATGTGAGTTGTTCTCCA 63  
11  
DB 641 AAGTGTGTAATGACAGAGTACGAGGACTGTATAAGACAATGTGTAATGACACAGA 700  
11  
QY 64 ATCCGCCAGAGGAGAGAGGCTTCTATGTTGACACAACAGTGTGATCTATAGTTGA 123  
11  
DB 701 TCCACACCAAAACGAGAGGCTTCATGTTTAAACCAAAATATCAATCTGTGGTTGA 760  
11  
QY 124 ATTCTGTACAGAACAAACACCAAGAGCTCCAAACAGCAAAATCAAAATATGC 181  
11  
DB 761 ATTCTGTACAGAACAAACACCAAGAGCTCCAAACAGCAAAATCAAAATATGC 818  
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

RESULT 5 AF001261 3317 bp mRNA MAM 19-NOV-1997  
LOCUS AF001261  
DEFINITION Bos taurus clone 1 endothelial adhesion molecule Lu-ECAM-1 mRNA.

complete cds.  
ACCESSION AF001261  
NID 92623762  
VERSION AF001261.1 GI:2623762  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 3317)  
AUTHORS Goodwin, A. and Paull, B.U.  
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an  
JOURNAL endothelial chloride channel  
REFERENCE 2 (bases 1 to 3317)  
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,  
Goodwin, A. and Paull, B.U.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,  
Ithaca, NY 14853, USA  
FEATURES  
source  
1. .3317  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/cell\_type="bovine endothelial cells"  
/clone="1"  
/tissue\_type="lung"  
63. .2780  
/note="endothelial adhesion molecule; chloride channel"  
/codon\_start=1  
/product="Lu-ECAM-1"  
/protein\_id="AAB86529.1"  
/db\_xref="PID:g2623763"  
/db\_xref="GI:2623763"  
/translation="NVLCALVILFLHLPLGKSSMNLINNGYDGVIAINPSPE  
DELINIKEMTEASTYFPAHTRRYRANSLILPMWKSKEFIPKOSYODAD  
VYANFLYLRKGDPTLQGRGCKRYTHFNPPLITPNPHYSGVRGVAHMAHL  
LMEFLYEVNPKQYAKESIMFPLSHVETCTEHTNTEANLNQKNCNGSTWDV  
LYNACFTFLPKSQTAKEIMFPLSHVETCTEHTNTEANLNQKNCNGSTWDV  
INNSVDFONTSPTENMPPTHPFSLKSKORVCLVLRKSGMSMEDRLFQNDAAE  
LVYIOYKSGSLGVTGTFDSVAEIOHNLRIEDNVYQKRTAKLPQVANGTSICGL  
KAGFOALIHSDOSTSGSEIILLDGEDNEINSGSEIDVKKSGALIHITIALGSPAARLE  
TISNMGCGYREFRANKDITGLITAFSISRSSTIQAALQLESKALIKGRVANGTV  
PVSTVGNDFEVVVTITQKPELVDDPKGKRYKTDREKDKINTPSARLOIPGIAET  
GWTYSLNLNHAASSQMLVYVTRASPITPIYATRAHNSQTAHIPSMTIYAQVSO  
GLPLVGLISVIAIETEDGHQVLELMDNAGADIVKNGISRYFTDYIGNRYSLK  
VHAQARNNTARLNLROPNKVLVYPGYVENGKILNLPREYVDDIADAKAIEDFSYLT  
SGGSFYSGAPPPGNNPSVPSKRTIDLEAKFEDYIQLSWAPGNVLKRGANSYII  
RISKSPMDROEDENATIVNTSLIPKEAGSEIEFEKPEHFEVNGTGYEIVQAIN  
EANDLSEVSHITVAIKFIFLPEDSVHDLGRTKISEIITLALGLPMIISVE"

BASE COUNT 1084 a 676 c 634 g 923 t

ORIGIN

Query Match 41.0%: Score 74.2; DB 3; Length 3317;  
Best Local Similarity 64.0%; Pred. No. 5.5e-09;  
Matches 112; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

7 ATGCACATTCATTAAGTATACAGAGCTCTATGAAAAAGATGTGAGTTGTTCTCCATC 66  
11  
DB 692 ATGCAGAGTATACACAGACAGGCTGTATGAGCAAAATGTACATCTCCCAAAAA 751  
11  
QY 67 CCGCCAGAGGAGAGGCTTCTATATGTTTGCACCAACATGTATCTATAGTTGAAT 126  
11  
DB 752 ATCCAGAGCTCCAAAGGATCATTAATGTTTATGCAAGTCTCATCTGTGACTGAAT 811  
11  
QY 127 CTGTACAGAACAAACCAAGAGCTCCAAACAGCAAAATCAAAATATGC 181  
11  
DB 812 TTGTACAGAACAAACCAACATACAGAGCTCCAAACCTACAAAACAAATATGC 866  
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```
RESULT 6
AF001262 3288 bp mRNA MAM 19-NOV-1997
LOCUS Bos taurus clone 2 endothelial adhesion molecule Lu-ECAM-1 mRNA,
DEFINITION complete cds.
ACCESSION AF001262
NID 92623764
VERSION AF001262.1 GI:2623764
KEYWORDS Bos taurus.
SOURCE Bos taurus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 3288)
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,
Goodwin, A. and Pauli, B.U.
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an
endothelial chloride channel
JOURNAL Unpublished
2 (bases 1 to 3288)
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,
Goodwin, A. and Pauli, B.U.
TITLE Direct Submission
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,
Ithaca, NY 14853, USA
FEATURES
source location/Qualifiers
1.3288
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_type="aortic endothelial cells"
/clone="2"
/tissue_type="lung"
59..2443
/note="endothelial adhesion molecule; chloride channel"
/codon_start=1
/product="Lu-ECAM-1"
/protein_id="AAB86530.1"
/db_xref="PID:92623765"
/db_xref="GI:2623765"
/translat="MVLCLNVLFLFLHLHPGKSSMNLNNGYDGIYAINPSVE
DKLEINEMVTEASTYLFHAKRKRVFRNVSILIPMTWKSSEFTIPKQSYQAD
VIVANPYKYGDPTLYGRGEGKGIHFPNPLLNPFHYSRGVFPVHNAHL
RMGIDEVAVDQPTYSRKNTEARCSHTHTGIVNVEFKCPGSCITSLCRDSOTG
LYEACFTLPKRSQAKESIMFSLHSVTECTEKTHTTEAPNLONKNCNCRSTWD
INNSVDFTSPMTMNPPTHTFSLSKORVVCYLDRSGMSAEDRLPQNOAAE
LYLQVIERGSLVGVTFDSVAEIONHLRITDNDVYOKITAKLPQVANGSTICGL
KAGPQALHSDOSTGSEITLLTDEDEINSCFEDVKRSGLIHTIALGPRSAKELE
TISNMGGRFRFANDITGLTNAFSRISRSSTIQAOLSKAKITGRRVNGTV
PDSVYGNDFVVTWITOKPEIVLQDPKRGKYSDFKEDLNTLSAKLQIPGIAET
GWTYSLLNNHASSQMLVTVTRASPPIPVIAVAHNSOHTAHPSPMIYAQVSQ
GLPYLGISVIAIIEFDGHOVTLLEMDNAGSDTVKNDGISRYETDYGGRGYSLK
VHAQARNNTARLROPNKRYLVYGVYENGKILNPPRPEVADLAKRIDFSRLT
SGGSFTVSGAPPNGNPSVFPSPSKITDLAKRKEDYIQLSWTAPGNVLDKXGSES"
BASE COUNT 1043 a 707 c 645 g 893 t
ORIGIN
Query Match 41.0%; Score 74.2; DB 3; Length 3288;
Best Local Similarity 64.0%; Pred. No. 5.5e-09;
Matches 112; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

```
RESULT 7
AF001263 2820 bp mRNA MAM 19-NOV-1997
LOCUS Bos taurus clone 3 endothelial adhesion molecule Lu-ECAM-1 mRNA,
DEFINITION complete cds.
ACCESSION AF001263
NID 92623766
VERSION AF001263.1 GI:2623766
KEYWORDS Bos taurus.
SOURCE Bos taurus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 2820)
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,
Goodwin, A. and Pauli, B.U.
TITLE Cloning and characterization of Lu-ECAM-1 suggest it is an
endothelial chloride channel
JOURNAL Unpublished
2 (bases 1 to 2820)
AUTHORS Elble, R.C., Widom, J., Gruber, A.D., Abdel-Ghany, M., Levine, R.,
Goodwin, A. and Pauli, B.U.
TITLE Direct Submission
JOURNAL Submitted (25-APR-1997) Vet. Pathology, Cornell U., VMC C4-146,
Ithaca, NY 14853, USA
FEATURES
source location/Qualifiers
1.2820
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="3"
/tissue_type="lung"
194..2656
/note="endothelial adhesion molecule; chloride channel"
/codon_start=1
/product="Lu-ECAM-1"
/protein_id="AAB86531.1"
/db_xref="PID:92623767"
/db_xref="GI:2623767"
/translat="MVLCLNVLFLFLHLHPGKSSMNLNNGYDGIYAINPSVE
DKLEINEMVTEASTYLFHAKRKRVFRNVSILIPMTWKSSEFTIPKQSYQAD
VIVANPYKYGDPTLYGRGEGKGIHFPNPLLNPFHYSRGVFPVHNAHL
RMGIDEVAVDQPTYSRKNTEARCSHTHTGIVNVEFKCPGSCITSLCRDSOTG
LYEACFTLPKRSQAKESIMFSLHSVTECTEKTHTTEAPNLONKNCNCRSTWD
INNSVDFTSPMTMNPPTHTFSLSKORVVCYLDRSGMSAEDRLPQNOAAE
LYLQVIERGSLVGVTFDSVAEIONHLRITDNDVYOKITAKLPQVANGSTICGL
KAGPQALHSDOSTGSEITLLTDEDEINSCFEDVKRSGLIHTIALGPRSAKELE
TISNMGGRFRFANDITGLTNAFSRISRSSTIQAOLSKAKITGRRVNGTV
PDSVYGNDFVVTWITOKPEIVLQDPKRGKYSDFKEDLNTLSAKLQIPGIAET
GWTYSLLNNHASSQMLVTVTRASPPIPVIAVAHNSOHTAHPSPMIYAQVSQ
GLPYLGISVIAIIEFDGHOVTLLEMDNAGSDTVKNDGISRYETDYGGRGYSLK
VHAQARNNTARLROPNKRYLVYGVYENGKILNPPRPEVADLAKRIDFSRLT
SGGSFTVSGAPPNGNPSVFPSPSKITDLAKRKEDYIQLSWTAPGNVLDKXGSSFP
MSRFSHQVAKVLEQLQHOHQ"
BASE COUNT 940 a 565 c 560 g 755 t
ORIGIN
Query Match 41.0%; Score 74.2; DB 3; Length 2820;
Best Local Similarity 64.0%; Pred. No. 5.7e-09;
Matches 112; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```



ORGANISM	Mus musculus
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
AUTHORS	1 (bases 1 to 3022)
TITLE	Gandhi, R., Eble, R.C., Gruber, A.D., Schreier, K.D., Ji, H.L., Fuller, C.M. and Pauli, B.U.
JOURNAL	Molecular and functional characterization of a calcium-sensitive chloride channel from mouse lung
MEDLINE	J. Biol. Chem. 273 (48), 32096-32101 (1998)
REFERENCE	2 (bases 1 to 3022)
AUTHORS	Gandhi, R., Eble, R.C., Gruber, A.D. and Pauli, B.U.
TITLE	Direct Submission
JOURNAL	Submitted (12-FEB-1998) Vet. Pathology, Cornell Univ., VMC C4-146, Ithaca, NY 14856, USA
FEATURES	location/Qualifiers
source	1..3022
gene	/organism="Mus musculus"
gene	/db_xref="taxon:10090"
gene	/tissue_type="whole lung"
gene	1..3022
gene	/gene="mCCLAL"
gene	18..2726
gene	/gene="mCCLAL"
gene	/note="calcium-activated chloride channel; Alternative chloride channel to CFTR that may influence severity of cystic fibrosis in different tissues; mCCL-1"
gene	/codon_start=1
gene	/product="calcium-sensitive chloride conductance protein-1"
gene	/protein_id="AAC79982.1"
gene	/db_xref="PID:g9325281"
gene	/db_xref="GI:3925281"
gene	/translation="MPGGLQVLEFLTLHLQNTSSWHLNSGEGVIAINSVEPEVERIPSEKEMVQASTYLEFEASQGRVYFNISLIVPMYKSEYELMPKRESYKADLVIDAPLHGGDDDDPTLOYGCGGGRGYIHETPNLITDNRIRYGRGRVHEMATHLMKGFDEYNDQPFYMSRKNTIETRESTRTGTVYHNCGRGCTYRACRDSKTLKTEPCFTIPDKTQTAGASIMFWQLNSVVECTEKNNHAEAPNLQNRKNSRWDTYKTSQDFONAPMRGTETEARPPPTFLKSRRVCLVLYDKSGMDKRLRMNAAELIYLTDIVKESWVGIVTFEDSAHIONLYLKIRSSDYDKITPDLQOASGGSICHGLQAGFAFISDOSTGSEIVLITDDEGNDIRSCFPAVSRAIHTITAGAPAAELETFELSDMTGLRFPANDINSILDAFRISTSGVSOALOLESKAFDVRAGWINGTVPRLDSIVGNDTFVITVMYKRPILILQDRGKRTTSDDQDKLRSARLQIPGAEFTG TWISITGTSQSLITMTVYTRARSPLEPLATAMSOSTRQYPSRMIYARVSQGLFVLGANTYALIEAHGQVLELDNGADTVANDIITYEYFDYDNGNGYSLKVVNOAGNKTRLSLRQKNKSLYIPGYENGKIVLNPDPVDEALATEDFERYVSGSSFTVSGAPDGDHARVPPSKVTLDEAEIGDYHLTWAPKVLVDNGRAHYITRMGSOHPDLQEDFNNAITVYNNSSSLIKPRAGSEYTFKPEFTKANGIOLYIALQDNESLITSEVSNIAQAVKLITSLSDSISALDDSDISATSMITWGLTVFNSTLN"
BASE COUNT	923 a 710 c 661 g 728 t
ORIGIN	
Query Match	39.3%; Score 71.2; DB 12; Length 3022;
Best Local Similarity	65.0%; Pred. No. 2.9e-08;
Matches 103: Conservative	0; Mismatches 53; Indels 0; Gaps 0;
Db	26 ACAGAGCTCTATGAAAAGATGTAGTGTGTCACATCCGCCAGAGGAGAAGCT 85
Db	663 ACACGGCTGATGAAACCCAAATGTCATTATTCACAGACAAATACAGACGGCTGGGCG 722
Db	86 TCTATTAATGTTTGCAACAACATGTTGATCTATAGTTGAATTCGTGACAAACAAACAC 145
Db	723 TCCCAATATGTCATGCAAAACCTTAATCTGTGGTTGATTTTGACAGAAAAAACCAC 782
Db	146 AACAAAGAGCTCCAAACAGCAAAATCAAAAATGC 181
Db	783 AATGCAGAGGCCCAAACTCAAAAACAAAATGTGC 818
RESULT 11	
LOCUS	AF052746 3471 bp mRNA ROD 08-APR-1999

DEFINITION	Mus musculus chloride channel cacc mRNA, complete cds.
ACCESSION	AF052746
NID	93560546
VERSION	AF052746.1 GI:3560546
KEYWORDS	
SOURCE	house mouse.
ORGANISM	Mus musculus.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Romfo, L., Musante, L., Cintl, R., Serl, M., Moran, O., Zegarra-Moran, O. and Galietta, L.J.V.
TITLE	Characterization of a murine gene homologous to the bovine Cacc chloride channel
JOURNAL	Gene 228 (1-2), 181-188 (1999)
MEDLINE	99173882
REFERENCE	2 (bases 1 to 3471)
AUTHORS	Romfo, L., Musante, L., Cintl, R., Moran, O., Serl, M. and Galietta, L.J.V.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAR-1998) 1db, di Genetica Molecolare, Istituto Giannina Gaslini, Genova I-16148, Italy
FEATURES	Location/Qualifiers
source	1..3471
CDS	/organism="Mus musculus" /db_xref="taxon:10090" /db_xref="dbest:AA245677" 47..2752 /codon_start=1 /product="chloride channel cacc" /protein_id="AAC35003.1" /db_xref="PID:93560547" /db_xref="GI:3560547" /translation="MVPGLQVLFLLTLHLQNTSSMYHLNSGVEGVIAINSPVE DIVRLPSIKEMVTOASTYLEFASOGRYAFRIHTPLILDNLRITGRGRVYVHEAHK VIALPILHOGDDPDTLOYGCGRGQYIHTPLILDNLRITGRGRVYVHEAHK RMGFEDRYNDQPFYMSRKNTIETRSTRITGVNYHNCRCNGCYTRACRDSKTR IYECFCITPMRLOTAGASIMFQNLNSVEVETCTKNHAEPLNQNKNCRSTWDVYR KTSQDFONAPMRGTLEAPPEPTSLKSRNVGLVLDKSGMKEDLIMNOAAEEL YLTDIVKESWQGVLFEDSAHIONLYLIKIRSSDYOKITANLPQASGGSICHGLD AGFAITSSDOSTGSEVILVLTDEGNDIRSCFPAVSRAIHTIALGPEARELEET LDMTGLRFPANQDINSILDAFRRISSTSGSVSOALQLFKAADVAGAWINGTVR LSTVGVNDTFVITWYKRPETIIQDKPKKTYTSDODKLNRSARLQIPETRETCT TWTSITGTNSQILTTWTVTRASPPIEPLATIAHMSOSTAQPSRMIVARVSGFT PVLGANTVLAIEAHGQVLTLELDNGAGDTKNDGIYTRFEDYHGNRSYSLKVR QAGNKTRLSLRQKNSLYIPGYENGIVLNPDPVDEAIATVEDFNRTVSGSGSICH FTVSGAPDGHARVFPSPKVTDLAEIGIYHLTWVAPKVLQNGVAHRYITMSG HPDLQEDFNHARTVYNNSSILPKPKASKEFKPEPTFKIANGVLIYALQADNEAS TSEVSNIAQAVKLTLSEDSIALDSDISALDICTMIGLVTFENSILN"
BASE COUNT	1096 a 777 c 736 g 862 t
ORIGIN	
Query Match	39.3%; Score 71.2; DB 12; Length 3471;
Best Local Similarity	66.0%; Pred. No. 2.9e-08;
Matches 103; Conservative	0; Mismatches 53; Indels 0; Gaps 0;
QY	26 ACAGAGCTCTAAGAAAAAGATGATGTTGTTCCCAATCCCGCAGCGAGAAGGCT 85
DB	632 ACAGCGCTGTATGAAGCCCAATATGTCATTATCCAGACAAATAATACAGCGGCTGGGCC 751
QY	86 TCTATTAATGTTTGCACAACAATGTTGATCTCTATAGTTGAATTTCTGTACAGAACAAACCC 145
DB	752 TCCATAATGTTTCATGCACAAACCTTAATCTGTGGTTGATTTTGCACAGAAAAAACCCAC 811
QY	146 AACAAAGAAGCTCCCAACAGCAAAAATCAAAAATGC 181
DB	812 AATGCAGAAAGCCCAAACTCAAAAACAAAATGTGC 847
RESULT 12	
BT036445	2984 bp mRNA MAM 09-FEB-1996
LOCUS	

	DEFINITION	Bos taurus calcium-activated chloride channel mRNA, complete cds.
ACCESSION	U36445	
NID	G1184065	
VERSION	036445.1	GI:1184065
KEYWORDS		
SOURCE	COW.	
ORGANISM	Bos taurus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos	
AUTHORS	1 (bases 1 to 2984) Cunningham,S.A., Awaaya,M.S., Buhlen,J.K., Ismailov,I.I., Arrate,M.P., Bendley,B.K., Benos,D.J. and Fuller,C.M.	
TITLE	Cloning of an epithelial chloride channel from bovine trachea	
JOURNAL	J. Biol. Chem. 270 (52), 31016-31026 (1995)	
MEDLINE	96125078	
REFERENCE	2 (bases 1 to 2984) Benos,D.J. Direct Submission Submitted (15-SEP-1995) Dale J. Benos, Physiology & Biophysics, University of Alabama at Birmingham, BHSB 706, Birmingham, AL 35294, USA	
FEATURES	Location/Qualifiers	
source	1..2984 /organism="Bos taurus" /db_xref="taxon:9913" /cell_type="epithelial cells" /tissue_type="trachea" 19..2750 /note="epithelial chloride channel" /codon_start=1 /product="calcium-activated chloride channel" /protein_id="AAC48511.1" /db_xref="pid:g1184066" /db_xref="gi:1184066"	
CDS	/translation="MPSRLVTLFELTLLHPGKMSWNLINNGDGVIAINPSVE DELIONIKENTASTYLFAHAKRRYRNVSLIPNTMSKEYLAMPQESYDAE VIANYPLKHGDPTITLOGRGCGEGCOIHETPFLNLLNIPIYSGRAFEEMAH RMQIDEVDGDPFYISRNLTJATRCSTHTGTGNVLYKGGSCIRPCRDSDGL YEKCFIPEKSQTARIESIMFMOSLSHATERCTEKTHVAEPNLOMKCKSTMVYI MLTDIFONTSPMTENMPPOPTFLSKRSORVCILDKSSGESDEBRTPMNDAEL FLIOIEKGSIVGWTFDSVAIERNNLRKIIDDNYENITANDLPQAMNGSTICRGK AGCAQAIISOOSTSGSEIILLFDGENEHSCIEEVKOSGYIHTIALGPSAAELEI LSDMTGHRFYANKDINGLNAPFSRISRSRSGSIIGQIIQLSKALATEKRWNGIVP VDSITGNDTFEVVTWKTKRPETLLDDPGKKRYKTSDEKEDKLHISARLIPGIAETG TWYSLLNNHASPOLLVITVRASPTPPPYATAHNSOMTAHYSPVIVAQVSOGQ FLPYGLNVAIAIETEDGHQVTELEMDNAGADYKNGCIYSRETFDRGGRSLAVY HAERNNRTARLSLRQPNOKRALIPGYENGKLIQLMPREPREDKLARKETIDERSLRS GAGSTVYGAPPNGHPSPVLNPKNKIIDLEAFKEDHLQSLSTPAVNLDLGRANSITIRI SKRTIDLOKDFDNALTYNTSLRKPRAGSDENFERKPEPFLENSTNYLIAVINERA NLSREVSNIAQAKETIIMPEDSVPALGTGISAINLAIFMALNISIV"	
BASE COUNT	1027 a 583 c 561 g 813 t	
ORIGIN		
Query Match	37.5%; Score 67.8; DB 3; Length 2984;	
Best Local Similarity	61.7%; Pred. No. 1.9e-07;	
Matches 108; Conservative	0; Mismatches 67; Indels 0; Gaps 0;	
Oy	7 ATGCACATTCAATTAAGTACAGGACTCTTAGAAAAAAGAGATGAGTTGTTCCCATC 66 	
Dd	645 ATGCACAGCTGACTCACACAGGAGGCTGTATGAGGCCAAAATGTACATTCAATCCCAGAAA 704 	
Oy	67 CCGCCAGAGGGAAGGCTTCATATGTTTGCCACACATGTGATTCATATGTTGAATT 126 	
Dd	705 ATCCCGAGCTGCAGAAGGATCATCATGTTATGCAAAGTCCATTCCTTGTGAGCTAATT 764 	
Oy	127 CTGTACAGAACCAACCACAACAGACTCCAAACAAGAAATCAAAATGC 181 	
Dd	765 TTGTACAGAAAAACACAAATGTGGAACTCCAAACTTACAAAACAAATAGTGG 819 	
RESULT	13	

LOCUS	AB026833	3604 bp	mRNA	PR	26-MAY-1999
DEFINITION	Homo sapiens mRNA for chloride channel protein, complete cds.				
ACCESSION	AB026833				
NID	48887600				
VERSION	AB026833.1	GI:4887600			
KEYWORDS	chloride channel protein.				
SOURCE	Homo sapiens adult corneal epithelium cDNA to mRNA, clone_l1b:lambda Zap clone:lambda 13802.				
ORGANISM	Homo sapiens				
REFERENCE	1 (sites)				
AUTHORS	Itch, R., Kawamoto, S., Kinoshita, S., Kawasaki, S. and Okubo, K.				
TITLE	Isolation and characterization of chloride channel in human corneal epithelium				
JOURNAL	Unpublished (1999)				
REFERENCE	2 (bases 1 to 3604)				
AUTHORS	Itch, R., Kawamoto, S. and Okubo, K.				
TITLE	Direct Submission				
JOURNAL	Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases. Reiko Itch, Institute for Molecular and Cellular Biology, Osaka University, Yamada-oka 1-3, Suita, Osaka 565-0871, Japan (E-mail:reiko@molc.biol.osaka-u.ac.jp, Tel:81-6-6879-7992, Fax:81-6-6877-1922)				
FEATURES	Location/Qualifiers				
source	1..3604				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="1"				
	/clone="lambda 13802"				
	/clone_lib="lambda Zap"				
	/dev_stage="adult"				
	/map="jg32"				
	/tissue_type="corneal epithelium"				
	42..2873				
	/codon_start=1				
	/product="chloride channel protein"				
	/protein_id="BA077810.1"				
	/db_xref="pid:d1041578"				
	/db_xref="pid:g4887601"				
	/db_xref="gi:4887601"				
	/translation="MTQRSIAGPICLNKFTVTLVALSSELPFGAGVQLODNGYNGLL /IAINOVPEKNLISINIKEMITEASFIFLNATKRVPRFRNKILIPNTNLTAGVSGSR OYSERKANYITVDYGAHGDPPYLYRGCGEKRTIHFPTNLTNLTAGVSGSR IVYHEWHLRMGVDEYINDRPFYINGNOQIKVRCSSDTIGFVCEKCGPOENCI STFKRGCFITNSQNTAFASIMFQSSLYVEFCNASTHQAENPQNLQNMCSLRSM DVTTSADPHHSPFNGNELPPPEFSLVQAGKYCVLDDYSKMAEADRLQLQQA AEYLQIYVEIFTFNGIASPFGSGEIRALQINSNDRLKLVGLYPTVSAKTDISI CSGLKRGFEVERLNGKATGASVMTITVSGDGLKNGCLPTVLSGSTHIALSSAA PNLETSRLTGLGKTFEPVDSNSMIDAFSLISGCGDIFQOHOLESTEENKPHH QLNLTIVNTVNGNDIMFLVTLQASGPEPILIFDPDGKTYTNNFTNLPTRTSLMT GVAKPGHWYTLNNTHSLQALKVTVSRASNSVPRATVAYVERSLIFHPHWI YAVKQGEYPLNAVTATVATEPEEDGPVTLRLDDGAGADVIKNGDITSRFFSPAN GRSLTVHNHSPSITSPAHSPGSHAMVPGVTNNGHINQMAKPSVGRNEEEKNG FSLKRSKSLQNTQDDPNNALVNTSKRMPQAGIRETFESQJISTGPHOPNGEET HEHSRIYVALRADKNSLDQSAVNTQAPLIFPPNSDPVARDYILKGVLLAMGLIGL IICLIIVYTHHTLSRRKKRADKENGKTKLL"				
BASE COUNT	1113 a	740 c	732 g	1017 t	2 others
ORIGIN					
Query Match	32.5%	Score 58.8	DB 9	Length 3604	
Best Local Similarity	62.0%	Pred. No. 2.7e-05			
Matches	93	Conservative	0	Mismatches 57	Indels 0
					Gaps 0
QY	32	CCTCATGAAAAAGATGATGAGTTGTCTCCATCCGCCACAGCAGGAAGCCTTCATATA	91		
DB	702	CTTTTAAAGAGATGACCTTATCTACAAATAGCACCCTAAATGCACTGCATTCATA	761		
QY	92	ATGTGGACAACAGTTCATCTATATCTGAATTCGTACAGAACCAACACACACAAA	151		







GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:35:28 ; Search time 3268.17 Seconds

(without alignments)  
281.231 Million cell updates/sec

Title: US-09-049-696-7

Perfect score: 289

Sequence: 1 GAAATATCCACTGATGAT.....ACCTTGAGAGTAAGGATTA 289

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBdb1:\*  
1: gb\_da1:\*  
2: gb\_da2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pa:\*  
6: gb\_ph:\*  
7: gb\_pi:\*  
8: gb\_pl:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: em\_fun:\*  
19: em\_hcg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pa:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sy:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hcg1:\*  
35: gb\_hcg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_da1:\*  
39: em\_da2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	238	82.4	3311	11	AF039400	Homo sapi

2	238	82.4	2826	42	AF127036	AF127036 Homo sapi
3	114	39.4	35278	11	AF039401	AF039401 Homo sapi
4	35	12.1	2937	12	AB017156	AB017156 Mus muscu
5	19	6.6	175825	9	HS106H8	Z97195 Human DNA s
6	19	6.6	169526	35	AC007081	AC007081 Drosophi1
7	18	6.2	136222	11	AC004099	AC004099 Homo sapi
8	18	6.2	1386	12	RND36785	U36785 Rattus norv
9	18	6.2	3804	12	RRRAGRF	X67241 R.rattus MR
10	18	6.2	462	14	G09420	G09420 human STS C
11	18	6.2	190000	35	AC006427	AC006427 Homo sapi
12	18	6.2	56479	42	AC006034	AC006034 Homo sapi
13	17	5.9	208430	1	BSQB0005	Z99108 Bacillus su
14	17	5.9	9444	1	D78508	D78508 Bacillus su
15	17	5.9	6349	3	BOYNEBXX1A	L14655 Bos taurus
16	17	5.9	2534	3	FC027351	U27351 Felis catus
17	17	5.9	19680	6	POT3111G	X17255 Bacteriophag
18	17	5.9	503	7	AB020483	AB020483 Vigna ung
19	17	5.9	34595	7	SPAC27E2	Z98978 S.pombe chr
20	17	5.9	106329	8	ATAC004411	AC004411 Arabidops
21	17	5.9	1408	8	FX063550	U63550 Fragaria x
22	17	5.9	5703	9	AB011150	AB011150 Homo sapi
23	17	5.9	4461	9	D87452	D87452 Human mRNA
24	17	5.9	198161	9	HS232K4	AL021938 Homo sapi
25	17	5.9	132981	9	HS82J11	Z83850 Human DNA s
26	17	5.9	10524	9	HDMHT21	D26350 Human mRNA
27	17	5.9	137947	10	HS4514	AL023581 Human DNA
28	17	5.9	100958	10	HS691N24	AL031672 Human DNA
29	17	5.9	94255	11	AC003090	AC003090 Homo sapi
30	17	5.9	100452	11	AC004499	AC004499 Homo sapi
31	17	5.9	132867	11	AC004601	AC004601 Human 11p
32	17	5.9	161652	11	AC004774	AC004774 Homo sapi
33	17	5.9	148915	11	AC005030	AC005030 Homo sapi
34	17	5.9	143865	11	AC005684	AC005684 Homo sapi
35	17	5.9	173717	11	AC005695	AC005695 Homo sapi
36	17	5.9	166701	11	AC005747	AC005747 Homo sapi
37	17	5.9	191111	11	AC006080	AC006080 Homo sapi
38	17	5.9	245439	12	MMHC135G15	AF050157 Mus muscu
39	17	5.9	5609	12	RATN1A	M96374 Rattus norv
40	17	5.9	155093	35	AC005948	AC005948 Homo sapi
41	17	5.9	145449	35	AC005949	AC005949 Homo sapi
42	17	5.9	168282	35	AC007322	AC007322 Homo sapi
43	17	5.9	262196	35	AC007513	AC007513 Homo sapi
44	17	5.9	167713	35	AC007778	AC007778 Homo sapi
45	17	5.9	16033	36	CEZK795	Z83246 Caenorhabdi

## ALIGNMENTS

RESULT 1	AF039400	3311 bp	mRNA	PRI	15-DEC-1998
LOCUS	AF039400				
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hCCAL) mRNA,				
complete cds.					
ACCESSION	AF039400				
NID	94009457				
VERSION	AF039400.1	GI:4009457			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	Gruber/A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.				
JOURNAL	Genomic cloning, molecular characterization, and functional				
MEDLINE	analysis of human CCLAL, the first human member of the family of				
REFERENCE	Ca2+-activated Cl- channel proteins				
AUTHORS	Genomics 54 (2), 200-214 (1998)				
TITLE	99047526				
	2 (bases 1 to 3311)				
	Gruber A.D., Elble, R. and Pauli, B.U.				
	Direct Submission				

**JOURNAL**  
Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

```

FEATURES
source      Location/Qualifiers
1. .3311   /org/ncim="HOMO sar

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p22-p31"
/tissue_type="small intestine"
1..3311
/gene="NCLCAL"
352..3096
/22..190191"
CDS
gene

```

BASE COUNT	1028 a	692 c	742 g	843
ORIGIN				

Query Match	82.4%;	Score 238;	DB 11;	Length 3311;
Best Local Similarity	99.7%;	Pred. No. 1.4e-123;		
Matches 288;	Conservative	0;	Mismatches 1;	Indels 0;

OY	1	GAATATCCAACTGATGATCTGAAATTTGGCTGCTGACGAAATGAGGAAAGAACACATCT	1604
Db	1545	GAATATCCAACTGATGATCTGAAATTTGGCTGCTGACGAAATGAGGAAAGAACACATCT	1604
OY	61	AAGTGGTGGTTTACGAGAGTCAAAACAAAGTNGTGCCATCATCCACAGTGGCTTTGGG	120
Db	1605	AAGTGGTGGTTTACGAGAGTCAAAACAAAGTGGTGCCATCATCCACAGTGGCTTTGGG	1664
OY	121	GCCCTGCGAGCTCAAGAACTAGAGAGAGTGTCCAAAATAGACAGAGTTTACAGACATA	180
Db	1665	GCCCTGCGAGCTCAAGAACTAGAGAGAGTGTCCAAAATAGACAGAGTTTACAGACATA	1724
OY	181	TGCTTCGATCAAGTTCASAAACAATGGCCCATTTGCTTTGGGGGCCCTTTCATCAGG	240
Db	1725	TGCTTCGATCAAGTTCASAAACAATGGCCCATTTGCTTTGGGGGCCCTTTCATCAGG	1784
OY	241	AAATGAGCTGTCTCTCAGCGCTTCATCCAGCTTGAGATAAGGAAATTA	289
Db	1785	AAATGAGCTGTCTCTCAGCGCTTCATCCAGCTTGAGATAAGGAAATTA	1833

RESULT	2		
AF127036			
LOCUS	2826 bp	mRNA	PRI
DEFINITION	Homo sapiens calcium-activated chloride channel protein 1 (CaCCl1)		16-APR-1999
ACCESSION	AF127036		
NID	94585468		
VERSION	AF127036.1	GI:4585468	

**KEYWORDS**  
**SOURCE**  
**ORGANISM**

SOURCE	ORGANISM
human.	human.
Homo sapiens	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	
1 (bases 1 to 2826)	
Agnew, X., Vernat, T. and Choussou, J.-M.	
Cloning of three human homologs of bovine epithelial chloride	

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2826)
AUTHORS	Agnel, M. and Culuscau, J.-M.
TITLE	Direct Submission
JOURNAL	Submitted (09-FEB-1999) Groupe Genomiche, SYNTHELABO, 10 Rue des Carrières, Rueil-Malmaison 92300, France
FEATURES	Location/Qualifiers
SOURCE	1..2826

BASE COUNT	875	a	623	c	632	g	696	t
ORIGIN								

Query Match	82.4%	Score 238;	DB 42;	Length 2826;
Best Local Similarity	99.7%	Pred. No. 1.4e-123;		
Matches 288;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0.

QY	1	GAATATCCAACTGATGGATGTGAAATTGTCTGTCTGACGGATGGGGAAGAACACATAT	60
Db	1198	GAATATCCAACTGATGGATGTGAAATTGTCTGTCTGACGGATGGGGAAGAACACATAT	1257
QY	61	AAGTGGGTGCTTTAACGAGCTCAACGAATAATNTGGCATCATCACAGCTGGCTTTGGG	120
Db	1258	AAGTGGGTGCTTTAAACGAGCTCAACGAATGGGTGGCATCATCACAGCTGGCTTTGGG	1317
QY	121	GCCCTCTCAGCTCAGAACAATCAGAGAGCTGTCCAAAATACAGAGGTTTACAGACATA	180
Db	1318	GCCCTCTCAGCTCAGAACAATCAGAGAGCTGTCCAAAATACAGAGGTTTACAGACATA	1377
QY	181	TGCTTCAGATCAAGTTCAGAACAAATGGCTCATATGATGCTTTGGGGCCCTTTCATCAGG	240
Db	1378	TGCTTCAGATCAAGTTCAGAACAAATGGCTCATATGATGCTTTGGGGCCCTTTCATCAGG	1433
QY	241	AAATGAGAGCTGTCTCTCAGGCTCCATCCACAGCTTGAGAGTAAAGGATTA	289

DB 1438 AAGTGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTA 1486

RESULT 3  
AF039401  
LOCUS AF039401 35278 bp DNA PRI 15-DEC-1998  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene, complete cds.  
ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
SOURCE  
1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/rpt\_family="LI transposable element"  
1587..1596  
/rpt\_type="dispersed"  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,6996..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,6996..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PIR:94009460"  
/db\_xref="GI:4009460"  
/translation="MGPFKSSYFIIILHLLEGALSNSLIQLNNNGEIVAVADPNP  
EDETLIQIKDVTQASLYLEFATGRRTFFKVVALLPEPTKTKADYRPRKLETKNA  
DVLVAESTPDKNDDEPTTEQMGNGCKGERIHILTPFIAKGKLAEGPQGAHVHMAH  
LWMGVDEYNNDKEFYLSNGRIQAVRCSGITGVNVKCOGSGCYTKRGTENKVTGL  
YKSGCEFLVQSTREKASIMFAOHVDSIYEFTEDNNHNEAPNRKONKCNLSTWEVI  
RDSBDFKKTTPMTIOPNPPTFSLQIGRIYCVLADKSGSNATGNRLNRINAGGLFL

107FELISWGMVTFEDSAHYOSELIQINSGBDPTLAKRLPAAASGGSITCSGLRSA  
FTVIRKRPPTDGESEIVLTDGEDNTISCFPNVYKOSGALHTFTVALGPRSAODELSK  
MTGGQTYASQVQVONNGLIDAFGALSSGNGAVSORSIQLESKGLTQNSQMMNGTVIV  
DSIVGKDFLITWTPPOILLNDPSGKOGGFEVDNFKMAVLIQIGIAKVGTMKY  
SLQASGDTLTLTISRASNATLPPITVTSKKNKDSKPSPLVYANROGASPLRA  
SVTALIESNKGRTVITLIDNGAGADATKDDGVYSRYPTTDTNGRSVYKRALGAVN  
AARRVITPOGSGALYIPGVTINDEIQWNPRRPEINKDDVQKQVCFSTSSGSPYAS  
DVPNAPIDLPFGQITDLNMEIHGGSLLINTWTAPEGDDIDGHIAKTIITISTSIDL  
LRDNFESLQVNTALALIPKEANSEVLFKPEPNTFEFGDILFAIQADVNDLKSEI  
SNIAVSLFIPPTPETPSPETSAPCPNIHINSTIPGIIHLIMMMWIGELQLSIA  
6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 39.4%; Score 114; DB 11; Length 35278;  
Best Local Similarity 99.4%; Pred. No. 8.4e-54;  
Matches 164; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GAAATTCGAAGTGAATGATGTAATTTGCTGTCGAGTGGGGAAGACACTAT 60  
DB 22264 GAAATTCGAAGTGAATGATGTAATTTGCTGTCGAGTGGGGAAGACACTAT 22323  
QY 61 AAGTGGTCTTAAAGAGGTCGAAACAAAGTNGCCATCATCCACACAGTGCCTTTGGG 120  
DB 22324 AAGTGGTCTTAAAGAGGTCGAAACAAAGTNGCCATCATCCACACAGTGCCTTTGGG 22383  
QY 121 GCCCTCTGAGCTCAAGACTGAGAGGCTGTCGAAATGAGAGG 165  
DB 22384 GCCCTCTGAGCTCAAGACTGAGAGGCTGTCGAAATGAGAGG 22428  
RESULT 4  
AB017156

LOCUS	AB017156	2937 bp	mRNA	ROD	04-MAR-1999
DEFINITION	Mus musculus gob-5 mRNA, complete cds.				
ACCESSION	AB017156				
NID	93721911				
KEYWORDS	AB017156.1	GI:3721911			
ORGANISM	Gob-5.				
REFERENCE	Mus musculus adult intestine goblet cell cDNA to mRNA.				
AUTHORS	Komiyama, T., Tanigawa, Y. and Hirohashi, S.				
TITLE	Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice				
JOURNAL	Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)				
MEDLINE	99160866				
REFERENCE	2 (bases 1 to 2937)				
AUTHORS	Komiyama, T., Tanigawa, Y. and Hirohashi, S.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru Tokuda, Tsukuba, Ibaraki 300-2635, Japan				
FEATURES	<p>Location/Qualifiers</p> <p>1..2937</p> <p>/organism="Mus musculus"</p> <p>/db_xref="taxon:10090"</p> <p>/cell_type="goblet cell"</p> <p>/dev_stage="adult"</p> <p>/tissue_type="intestine"</p> <p>15..2756</p> <p>/gene="gob-5"</p> <p>15..2756</p> <p>/gene="gob-5"</p> <p>/codon_start=1</p> <p>/protein_id="BAA33743.1"</p> <p>/db_xref="pid:d1034712"</p> <p>/db_xref="gi:3721912"</p> <p>/translation="MESLSPVFLIIHLLEGVLSSESLQLNNNGYEGYVLAIDHPV EDFALQHKDMVTOASPYLFEATGKRRYFKNVAILIPESKARPETRPLETFKNA DVLVETSPLDGDEPYTEHIGACGEGIRIHLPFLAGKLTQYGGDDTFVEMAH FMGVPENYNNDEKRYLSKGPQAVRCSAAITGKNQVRCOGGSCINDGKCIDRVG LRKDNVPEYPPDHONKASIMNOMNINVECEKRNQNOAPNDONORCMLRSTWVE IOESDFKQTFPMIAQPPAPTSLLQIQORIVCIYLDKSGMLDNRNRNQSRLF LQAYEOGSGWGVMTFDSAAYOSEKQNSGADBDILIKHLPIVSAAGTSICGLPT KMTGKIKRYPDQGESEIVLITDGEDNTISSCPVLVQSGAIIHTYVALGPAAKKEEPLS ATTGVLKQYSSDOVONNGVLDAFALISSGNAIAQHSIQLESRGVNLNOMWANGSYI VNSVSGKDTLFLITWTTHPTPIFIMDPSSVEONGFIIDPTTRVALOYGTGKGFVK VSSQASQRTLLTIVTSRAASATLPPIVTPVYNNKTKGFSPSVYASIRGASPIILR ASTYALIESVNNKRTVTLITLNDGAGADATKNGVSRREFADANGRSVATMALGGV TSRQRAAPKRNKRAMITDMITGIVTGLKASITOGCNLYNLWTAPGDYDHGRASNTIMRSIYDV VPAAPAIPLDFPCCQITDLKASITOGCNLYNLWTAPGDYDHGRASNTIMRSIYDV LRNHFMTSLQVNTGLIPKEASSELETFEEELGNGTFGKNTDIFLAIQAVDKSNKSEEL SNIAKRSVFIPAOEPDIPEDSPPCDDISINSLTIFGLHVLKIMKMLGEMOYTTGLH</p>				
gene					
CDS					
BASE COUNT	860 a	718 c	693 g	666 t	
ORIGIN					
Query Match	12.1%; Score 35; DB 12; Length 2937;				
Best Local Similarity	100.0%; Pred. No. 2.8e-09;				
Matches	35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	5 TATCCAACTGATGATCGATCGAATGTGCTGCTGAC 39				
Db	1215 TATCCAACTGATGATCGAATGTGCTGCTGAC 1249				
RESULT	5				
LOCUS	HS106H8 175825 bp DNA PRI 10-FEB-1999				
DEFINITION	Human DNA sequence from pBC 106H8 on chromosome 1q24. Contains				

```

ACCESSION      297195
NID            92578039
VERSION       297195.1
KEYWORDS      1924: Cpc; DNM; DYN3; DYNAMIN; PHOSPHATIDYLINOSITOL-GLYCAN; PIG-C.
SOURCE        human.
ORGANISM      Homo sapiens
AUTHORS       Enkaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 175825)
TITLE         Grafham/D.
JOURNAL       Submitted (14-OCT-1997) Chromosome 1 Project Group
              (http://www.sanger.ac.uk/HGP/Chr1/) Sanger Centre, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              cambridge@sanger.ac.uk
              On Oct 31, 1997 this sequence version replaced gi:2225976.
              IMPORTANT: This sequence is not the entire insert of clone 106H8.
              It may be shorter because we only sequence overlapping sections
              once, or longer because we arrange for a small overlap between
              neighbouring submissions.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variations annotated may not be found in the sequence submission
              corresponding to the overlapping clone as we submit sequences with
              only a small overlap as described above.
              This sequence was generated from part of bacterial clone contigs of
              human chromosome 1, constructed by the Sanger Centre chromosome 1
              mapping group. Further information can be found at
              http://www.sanger.ac.uk/HGP/Chr1/
              This sequence has been finished according to sequence map criteria
              as follows. An attempt is made to resolve all sequencing problems,
              such as compressions and repeats, but not necessarily within known
              annotated human repeat sequence elements (e.g. Alu). Where the
              sequence is ambiguous, there is an annotation using the 'unsure'
              feature key.
              The true right end of clone 125H23 is at 104 in this sequence. The
              true right end of clone 106H8 is at 175825.
              106H8 is from the library RPC11 constructed at the Roswell Park
              Cancer Institute by the group of Pieter de Jong.
              For further details see http://bacpac.med.buffalo.edu/.

FEATURES
    source
        1..175825
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /map="1q24"
            /clone_1lb="106H8"
            /clone_1lb="RPC11"
            /clone_1lb="2571"
            /note="AluX repeat: matches 302. .1 of consensus"
            /note="AluX repeat: matches 6. .63 of consensus"
            /note="MER39 repeat: matches 1. .299 of consensus"
            /note="AluY repeat: matches 1. .299 of consensus"
            /note="MER39 repeat: matches 57. .460 of consensus"
            /note="AluX repeat: matches 1. .302 of consensus"
            /note="AluX repeat: matches 1191. .126 of consensus"
            /note="MER42c repeat: matches 365. .1 of consensus"
            /note="MLT1 repeat: matches 296. .1 of consensus"
            /note="AluY repeat: matches 296. .1 of consensus"
            /note="AluX repeat: matches 300. .1 of consensus"
            /note="AluY repeat: matches 691. .484 of consensus"
            /note="L1r8 repeat: matches 691. .484 of consensus"

```

repeat_region	14970..15411	/note="LRR8 repeat: matches 406. .1 of consensus"
repeat_region	10310..16411	/note="3 copies of 34 mer 86 & conserved"
repeat_region	17515..17585	/note="MIR repeat: matches 90. .162 of consensus"
repeat_region	18775..19095	/note="AluB repeat: matches 1. .300 of consensus"
repeat_region	19194..19519	/note="AluSx repeat: matches 302. .1 of consensus"
repeat_region	20512..20636	/note="MIR2 repeat: matches 17. .146 of consensus"
repeat_region	20662..20967	/note="AluSx repeat: matches 301. .9 of consensus"
repeat_region	22214..22326	/note="MIR2 repeat: matches 146. .41 of consensus"
repeat_region	22361..22574	/note="MIR repeat: matches 12. .258 of consensus"
repeat_region	24516..24601	/note="MIR repeat: matches 109. .187 of consensus"
repeat_region	27128..27347	/note="MIR repeat: matches 259. .28 of consensus"
repeat_region	28162..28293	/note="FLAM_C repeat: matches 1. .133 of consensus"
repeat_region	28507..28622	/note="MIR2 repeat: matches 24. .142 of consensus"
repeat_region	30226..30287	/note="MIR repeat: matches 45. .108 of consensus"
repeat_region	30288..30584	/note="AluSx repeat: matches 297. .1 of consensus"
repeat_region	30585..30743	/note="MIR repeat: matches 101. .260 of consensus"
repeat_region	31440..31626	/note="MIR2 repeat: matches 189. .2 of consensus"
repeat_region	31743..31854	/note="MIR repeat: matches 18. .131 of consensus"
repeat_region	31870..32003	/note="MIR repeat: matches 251. .108 of consensus"
repeat_region	33084..33156	/note="MIR2 repeat: matches 62. .136 of consensus"
repeat_region	33180..33641	/note="LRR2 repeat: matches 449. .1 of consensus"
repeat_region	34747..34953	/note="MIR4B repeat: matches 221. .9 of consensus"
repeat_region	35115..35213	/note="MIR39 repeat: matches 516. .416 of consensus"
repeat_region	35296..35666	/note="MIR39 repeat: matches 460. .78 of consensus"
repeat_region	36591..36894	/note="AluSx repeat: matches 302. .1 of consensus"
repeat_region	38652..38823	/note="MIR repeat: matches 174. .1 of consensus"
repeat_region	39473..39570	/note="MIR4B repeat: matches 1. .93 of consensus"
repeat_region	39623..39726	/note="MIR4B repeat: matches 110. .208 of consensus"
repeat_region	39909..40087	/note="MIR4B repeat: matches 338. .502 of consensus"
repeat_region	40871..41162	/note="AluSg repeat: matches 7. .300 of consensus"
repeat_region	43088..43183	/note="MIR repeat: matches 99. .5 of consensus"
repeat_region	44364..44646	/note="AluB repeat: matches 1. .272 of consensus"
repeat_region	44736..44837	/note="3 copies of 34 mer 87 & conserved"
repeat_region	46169..46214	/note="MIR2 repeat: matches 100. .146 of consensus"
repeat_region	46501..46561	/note="AluJ repeat: matches 301. .241 of consensus"
repeat_region	46563..46764	/note="MIR repeat: matches 199. .1 of consensus"
repeat_region	47089..47416	/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region	48314..48537	/note="AluB repeat: matches 84. .302 of consensus"
repeat_region	49327..49471	/note="MIR2 repeat: matches 1. .146 of consensus"
repeat_region	49831..50119	/note="AluJ repeat: matches 301. .1 of consensus"
repeat_region	51640..51733	/note="MIR repeat: matches 260. .151 of consensus"
repeat_region	51763..52197	/note="MIR39 repeat: matches 516. .78 of consensus"
repeat_region	52248..52381	/note="MIR repeat: matches 158. .10 of consensus"
repeat_region	53816..53920	/note="MIR repeat: matches 83. .190 of consensus"
repeat_region	55448..55487	/note="MIR repeat: matches 20. .18 of consensus"
repeat_region	59212..59247	/note="18 copies of 2 mer 83 & conserved"
repeat_region	60066..60197	/note="FLAM_C repeat: matches 131. .1 of consensus"
repeat_region	60232..60535	/note="AluSg repeat: matches 301. .1 of consensus"
repeat_region	60770..60829	/note="AluSg repeat: matches 60. .1 of consensus"
repeat_region	61017..61395	/note="match 3' fetal heart ESTs W81520 W56365"
repeat_region	63217..63274	/note="29 copies of 2 mer 95 & conserved"
repeat_region	64641..64936	/note="AluSg repeat: matches 301. .7 of consensus"
repeat_region	66771..67069	/note="AluSx repeat: matches 5. .299 of consensus"
repeat_region	69198..69239	/note="MIR2 repeat: matches 96. .137 of consensus"
repeat_region	69646..70539	/note="LRR5 repeat: matches 900. .1 of consensus"
repeat_region	70391..71787	/note="MIR repeat: matches 5390. .3957 of consensus"
repeat_region	76178..76264	/note="MIR2 repeat: matches 60. .146 of consensus"
repeat_region	76660..77653	/note="match 5' ESTs W87669 H71966"
repeat_region	77525..77583	/note="MIR repeat: matches 142. .84 of consensus"
repeat_region	77673..77973	/note="AluSg repeat: matches 1. .300 of consensus"

ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 169526)  
 AUTHORS Ceoliker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazey,R.G.,  
 Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,  
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Katta,K., Kearney,L.,  
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclob,J.M., Park,S.,  
 Pfeiffer,B., Poon,L., Sequiera,A., Sethi,H., Snir,E.,  
 Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
 Rubin,G.M.  
 TITLE Sequencing of Drosophila melanogaster  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 169526)  
 AUTHORS Ceoliker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazey,R.G.,  
 Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,  
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Katta,K., Kearney,L.,  
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclob,J.M., Park,S.,  
 Pfeiffer,B., Poon,L., Sequiera,A., Sethi,H., Snir,E.,  
 Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
 Rubin,G.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-MAR-1999) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 COMMENT On Jun 1, 1999 this sequence version replaced g1:4803907.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [help@fruitfly.org](mailto:help@fruitfly.org). All contigs in this submission meet  
 the following cutoffs: length >= 400 bases, phrap computed error  
 rate <= 1/10.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 733: contig of 733 bp in length  
 \* 734 813: gap of unknown length  
 \* 814 41210: contig of 40397 bp in length  
 \* 41211 41290: gap of unknown length  
 \* 41291 169526: contig of 128236 bp in length.  
 \* Location/Qualifiers  
 \* 1. 169526  
 \* /organism="Drosophila melanogaster"  
 \* /strain="y: cn bw sp"  
 \* /db\_xref="taxon:7227"  
 \* /clone="BACR48023 (D498) RPCI-98 48.0.23"  
 \* /chromosome="3"  
 \* /clone\_11b="RPCI-98 (Roswell Park Cancer Institute  
 \* Drosophila melanogaster BAC library, partial EcoRI in  
 \* pBACE3.6"  
 \* /map="99A7-99B5"  
 \* BASE COUNT 47509 a 37630 c 37526 g 46700 t 161 others  
 \* ORIGIN  
 \* Query Match 6.6%; Score 19; DB 35; Length 169526;  
 \* Best Local Similarity 100.0%; Pred. No. 2.4;  
 \* Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 \* QY 213 TTGATGCTTTGGGCGCT 231  
 \* Db 81329 TTGATGCTTTGGGCGCT 81311

RESULT 7  
 AC004099  
 LOCUS  
 DEFINITION  
 ACCESSION  
 NID  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Plimates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 136222)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 17, clone Hc17A21K24  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 136222)  
 AUTHORS Birren,B., Fasmann,K., McKernan,K., Nusbaum,C., Richardson,P.,  
 Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckert,R.,  
 Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J.,  
 Depayre,E., Devon,K., Dewar,K., Duret,B., Etemad,S.,  
 Ferreira,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S.,  
 Geraghty,K., Gilman,T., Gray,D., Hagos,B., Harris,K.,  
 Horton,L., Howland,J.C., Hul,L., Jacotot,L., Linton,L.,  
 Mackenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrum,J.,  
 Molli,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T.,  
 Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D.,  
 Rollins,G., Rossello,R., Roy,A., Shyam,R., Soboo,S., Stillwell,J.,  
 Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A.,  
 Wagner,A., Wheeler,J., Wu,Y., Ye,W.J., Zemtseva,I., Zhao,J. and  
 Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-FEB-1998) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 3 (bases 1 to 136222)  
 AUTHORS Birren,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,  
 Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Boatn,C.,  
 Boutwell,C., Brown,A., Byrne,S., Cantu,C., Castle,A., Cerny,D.,  
 Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Duret,B.,  
 Etemad,S., Ferreira,P., Fitzgerald,W., Forrest,C., Funke,R.,  
 Gage,D., Gardyna,S., Gensheimer,S., Geraghty,K., Gilman,T.,  
 Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hul,L.,  
 Jacotot,L., Kann,L., Macdonald,P., Marquis,N., McEwan,P.,  
 McGurk,A., McKernan,K., Meldrum,J., Molli,M., Morris,W., Morrow,J.,  
 Mychaleckyj,J., Nachman,A., Naylor,J., O'Connor,T.,  
 Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A.,  
 Shyam,R., Stange-Thomann,N., Stillwell,J., Stone,C., Strickland,C.,  
 Subramanian,A., Toriella-Miller,I., Vassiliev,H., Vo,A.,  
 Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-MAY-1998) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 4 (bases 1 to 136222)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Boutwell,C.,  
 Brown,A., Castle,A., Cerny,J., Collins,D., Depayre,E., Devon,K.,  
 Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,  
 Donelan,L., Ferreira,P., Fitzgerald,W., Forrest,C., Funke,R.,  
 Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heatford,A.,  
 Heinen,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,  
 Karadas,A., Lechoczky,J., Macdonald,P., Marquis,N., McEwan,P.,  
 McGurk,A., McKernan,K., Meldrum,J., Molli,M., Morris,W., Morrow,J.,  
 Mychaleckyj,J., Naylor,J., Naylor,J., Naylor,J., O'Connor,T.,  
 O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,  
 Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,  
 Subramanian,A., Tesfaye,S., Tichovolsky,N., Toriella-Miller,I.,  
 Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,  
 Ye,W.J., Zhao,J. and Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-SEP-1998) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Sep 25, 1998 this sequence version replaced g1:3128780.  
 All repeats were identified using RepeatMasker; Smit, A.F.A. &

Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html.  
Location/Qualifiers

FEATURES  
source

```

1. .136222
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HCTR421K24"
/clone_11b="Research Genetics/Cal Tech CITB-HSP-C (plates
195-384)"
/map="17"
/chromosome="17"
257. .525
/rpt_family="AluDb"
637. .945
/rpt_family="AluDb"
complement(1015. .1189)
/rpt_family="AluSg"
complement(1190. .1487)
/rpt_family="AluSg"
complement(1488. .1624)
/rpt_family="AluSg"
complement(1626. .1667)
/rpt_family="(TAA)n"
complement(1668. .1944)
/rpt_family="AluSx"
complement(1958. .2074)
/rpt_family="L1MC4"
2101. .2380
/rpt_family="AluDo"
complement(2882. .3191)
/rpt_family="AluSg"
3285. .3502
/rpt_family="AluSp"
complement(3503. .3614)
/rpt_family="AluSc"
3619. .3673
/rpt_family="Alu"
complement(3692. .4000)
/rpt_family="AluSp"
complement(4002. .4323)
/rpt_family="AluY"
complement(4479. .4663)
/rpt_family="MIR"
complement(4796. .5076)
/rpt_family="AluDo"
5097. .5124
/rpt_family="AT_r1ch"
complement(5128. .5283)
/rpt_family="FRAM"
complement(5338. .5392)
/rpt_family="MIR"
5534. .5833
/rpt_family="AluSg"
5847. .5999
/rpt_family="FRAM"
6842. .7136
/rpt_family="AluSg1"
complement(7291. .7526)
/rpt_family="MTR1E"
complement(7555. .7621)
/rpt_family="MTR1D"
complement(7741. .7908)
/rpt_family="AluY"
7909. .8217
/rpt_family="AluY"
complement(8218. .8233)
/rpt_family="AluY"
complement(8266. .8382)
/rpt_family="AluY"
complement(8395. .8502)
/rpt_family="MTR1D"
complement(8572. .8724)
/rpt_family="AluDb"

```

```

repeat_region complement(9729. .9842)
/rpt_family="MER5B"
repeat_region complement(10333. .10478)
/rpt_family="MTR1F"
repeat_region complement(10627. .10706)
/rpt_family="L1MC5"
repeat_region complement(10707. .11009)
/rpt_family="AluSx"
repeat_region complement(11010. .11093)
/rpt_family="L1MC5"
repeat_region complement(11166. .11299)
/rpt_family="L1ME2"
repeat_region complement(11300. .11608)
/rpt_family="AluSx"
repeat_region complement(11609. .12072)
/rpt_family="L1ME2"
complement(12070)
/Note="Probably 28 As. Three clones read 28 As, one clone
reads 27 As."
repeat_region complement(12093. .12213)
/rpt_family="L1MC5"
repeat_region complement(12229. .12273)
/rpt_family="AT_r1ch"
repeat_region complement(12280. .12396)
/rpt_family="AluDb"
repeat_region complement(12414. .12520)
/rpt_family="L1MD3"
repeat_region complement(12814. .12814)
/rpt_family="AluDo"
repeat_region complement(13111. .13111)
/rpt_family="AluSx"
repeat_region complement(13112. .13455)
/rpt_family="L1MD3"
repeat_region complement(13596. .13596)
/rpt_family="MER21B"
repeat_region complement(13603. .13748)
/rpt_family="L1MC5"
repeat_region complement(13837. .14111)
/rpt_family="L1MC/D"
repeat_region complement(14112. .14284)
/rpt_family="AluSp"
repeat_region complement(14285. .14409)
/rpt_family="L1MC4"
repeat_region complement(14413. .14450)
/rpt_family="(CAA)n"
repeat_region complement(14451. .14733)
/rpt_family="AluY"
repeat_region complement(14736. .15025)
/rpt_family="AluSg"
repeat_region complement(15026. .15160)
/rpt_family="AluDb"
repeat_region complement(15503. .15804)
/rpt_family="AluSg"
repeat_region complement(16223. .16536)
/rpt_family="AluY"
repeat_region complement(16539. .16832)

```

Query Match 6.2%; Score 18; DB 11; Length 136222;  
Best Local Similarity 100.0%; Pred. NO. 9;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TGACAGGAGGTTACAGA 176  
DB 130510 TGACAGGAGGTTACAGA 130527

RESULT 8  
RNU36785 1386 bp mRNA ROD 12-MAR-1996  
LOCUS Rattus norvegicus putative pheromone receptor VNI mRNA, complete  
DEFINITION cds.  
ACCESSION U36785  
NID 91039469

```

VERSION      036785.1  GI:1039469
KEYWORDS
SOURCE
ORGANISM      Norway rat strain-Sprague-Dawley.
              Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Rattus.
REFERENCE
AUTHORS      Dulac, C. and Axel, R.
TITLE        1 (bases 1 to 1386)
              A novel family of genes encoding putative pheromone receptors in
              mammals
JOURNAL      Cell 83 (2), 195-206 (1995)
MEDLINE      96028094
REFERENCE
AUTHORS      Dulac, C. and Axel, R.
TITLE        2 (bases 1 to 1386)
              Direct Submission
              Submitted (20-SEP-1995) Catherine Dulac, Biochemistry and Molecular
              Biophysics, Howard Hughes Medical Institute Columbia University,
              701 West, 168th Street, New York, NY 10032, USA
FEATURES
SOURCE
CDS
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="vomeronasal organ"
/sex="female"
701.1217
/note="putative: seven transmembrane domain receptor"
/codon_start=1
/product="pheromone receptor VN1"
/protein_id="AAA92007.1"
/db_xref="PIR:G1039470"
/db_xref="GI:1039470"
/translation="MAKNKSLRYTDSNIRNFEAFIGVANSLLLFNFKLIGQ
RRLDLPGLSLINLNLMLTARLADPDSWGMGDDICKSLIYRFRGSLC
RSCSLSVLAQIILSPSSCLAFKRPSPHISCAISLYLMPFSSHLVITATP
LTTNPIHYTQWCSILIPMSYLMQSMFSLIARVFLISLANTSYRYVALCHRRK
TTHLOSTLSPKASPEQRATRSILIMLSLFVLMVSFDSIVCSSRTMYNDPISYD
FMWIRATVSPFVYTEKHIVNSMCKVKECLINP"
BASE COUNT   357 a      322 c      271 g      435 t      1 others
ORIGIN
Query Match      6.2%; Score 18; DB 12; Length 1386;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
DY 125 TCTGCAGCTCAAGAACTA 142
Db 1042 TCTGCAGCTCAAGAACTA 1059
RESULT 9
LOCUS      RRRASGRF      3804 bp      mRNA      ROD      03-DEC-1993
DEFINITION R.rattus mRNA for guanine nucleotide releasing factor for Ras p21.
ACCESSION  X67241 S40901
NID        957664
VERSION    X67241.1 GI:57664
KEYWORDS   guanine nucleotide releasing factor; releasing factor.
SOURCE     black rat.
ORGANISM   Rattus rattus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathii; Muridae; Murinae; Rattus.
REFERENCE
AUTHORS    Feig, L.A.
            Direct Submission
            Submitted (07-JUL-1992) L.A. Feig, Tufts Univ. School of Medicine,
            136 Harrison Ave., Boston MA 02111, USA
JOURNAL    2 (bases 1 to 3804)
            62 (bases 1 to 3804)
            Shou, C., Farnsworth, C.L., Neel, B.G. and Feig, L.A.
            Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing
            factor for Ras p21
            Nature 358 (6384), 351-354 (1992)
JOURNAL

```

MEDLINE 923505260  
 FEATURES  
 source  
 Location/Qualifiers  
 1. 3804  
 /organism="Rattus rattus"  
 /db\_xref="taxon:10117"  
 /dev\_stage="adult"  
 /tissue\_type="brain"  
 /clone\_lib="lambda gt10"  
 70. 3804  
 /codon\_start=1  
 /product="p140 RAS-GRF"  
 /protein\_id="CAA47666.1"  
 /db\_xref="pid:957665"  
 /db\_xref="GI:57665"  
 /db\_xref="SWISS-PROT:P28818"  
 /translation="MOKAIRNDGHVYSIGLLAORNGTRKGYLSKSSDNPKWQTKWF  
 ALLQNLFFYESDSSSRPSGLYLLBESITCKRMPSPRGSSKESQOHYFVNFNSD  
 SOSLELRDSDSCDEWVAALRAISKIKLATEHEALMOKYLIHLIOVETFEVAKMC  
 ROOLEGEVEIERLKAELIANLKDNRISDNOLVAEEDSDIKAKYQSLRWKIC  
 KRNKNIIODYIRSPHADSWMKRNQYVEFMLEEAAYQVQLILVNNFLPRLMASS  
 RRKMKNIIDYISIFNSERTIMEFHLQITGYGKARIKISPTLFIYPMQIPRYLITQ  
 KRPPIHDVSSIFNSERTIMEFHLQITGYGKARIKISPTLFIYPMQIPRYLITQ  
 EYRNHYSIOGLIHLKONDRDFEKLKQYEAKDCDEERLETFILYPMQIPRYLITQ  
 HELLAIRPHEHVERNSIDYAKSKLEIRSYMHDEYSETENIRKILAEIMTEGCEBIL  
 LDTSQTFRQGSILQYPMSEKSKIKNGRGSISLKEGEGOCLEFKHLILCTRGSGS  
 KLHTRNGVSLIDCTLLDPENMDDGKGQEVPHDLPFRIMWPKSDPSPFTYILVAS  
 ROEKTAANTSDIIQCVNDIRONGLMANNFENSKSVYTPQMKISQACDDVIDREKST  
 MNSCKVAQIRYASVERLLERLTLRLSLIDFLSLIRVTLASTIRVTDANYVLDKLSYKK  
 NPIPARSELLELFFSSSHNTKLLYGDAPKSPRASRSFSPPLATITSSPYRRKRLSL  
 NPIITGKALELLELFGCPDGDTYTNHISPISPGKTTLDSIKCVASSLRTPEEDLN  
 TLTDESGFRKPTSDILKESSDDQSDVDTEVSPPPKSFFRRRIITQEPFLFVNSGI  
 MTRCDSMGNSRSPLSATISAFIATAGANSPPANKELIYRMSLANTGQVSSDQRIKDE  
 FVIRARANTRYLVNLRHMVTRKHSODEFTDILLKRYVCEFLFEMHPDILLPOERKAA  
 NIKMTILQTEELTENVHSMILDELIMTEGVTCEPENSAMEIAEQLTLDLHLSKIPY  
 NEMKRTUOEGEENKDERKERTPIYIMKTRFRRHNHISNLIASEIINEEVSAASTIERKVA  
 EEFGQGMKMKADNRERTPIYIMKTRFRRHNHISNLIASEIINEEVSAASTIERKVA  
 DICGLAHNYNVALEITSSINSALFRKAKMIVYSKOTKSLFPKLOKIVSDGRFNL  
 RETLRNDPCVCPYLGWILDLAFLEGTENYTEDGYNVSKRMKIMSHIIRLEIROQO  
 TTYKIEPQPKVQYLVDETFVLDDESLEYASLRLEPKLP1"  
 238. 461  
 /note="Dh1 oncogene-like"  
 920. 1221  
 /note="S.cerevisiae CDC-25-like"  
 /note="S.cerevisiae CDC-25-like"  
 1015 a 1077 c 932 g 780 t  
 BASE COUNT  
 ORIGIN  
 Query Match 6.2%; Score 18; DB 12; Length 3804;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 138 AACTAGAGAGAGCTGTCCA 155  
 |||  
 |||  
 Db 1313 AACTAGAGAGAGCTGTCCA 1330  
 RESULT 10  
 LOCUS G09420/c 462 bp DNA STS 14-AUG-1995  
 DEFINITION human STS CHLC.GCT1C11.P9679 clone GCT1C11.  
 G09420  
 NID 9941269  
 G09420.1 GI:941269  
 VERSION STS sequence; primer: sequence tagged site.  
 KEYWORDS human vector-pCp1 host-e.coli dut-tung+ (DH10B) Marker Selected  
 SOURCE genomic DNA prepared from XY individual of French nationality.  
 ORGANISM Homo sapiens  
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;  
 Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates  
 Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 462)  
 Murray, J., Sheffield, V., Weber, J.L., Duyk, G. and Buetow, K.H.  
 Cooperative Human Linkage Center  
 Unpublished (1995)



## COMMENT

Synonyms: GCTIC11, CHLC, GCTIC11.19678  
Contact: Dr. Jeffrey C. Murray

## Unit

The University of Iowa  
Department of Pediatrics, Iowa City, IA 52242, USA  
Tel: (319) 356-3508  
Fax: (319) 356-3347  
Email: jeff-murray@uiowa.edu

Primer A: TGACTAGAGGAAAAAGCTTGG  
Primer B: TTAGTCTGACACGGGAAAG  
STS size: 122

## PCR Profile:

denature: 30 seconds at 94 degrees C  
annealing: 75 seconds at 55 degrees C  
extension: 15 seconds at 72 degrees C  
PCR cycles: 27  
extension: 6 minutes at 72 degrees C

## Protocol:

Template: 30ng genomic DNA  
Primer: each 1.5 pmole  
dNTPs: each 200 uM  
Tag Polymerase: 0.3 units  
Total Vol: 10 uL

## Buffer:

MgCl2: 1.5mM  
KCl: 50mM  
Tris: 10mM  
pH: 8.3

## FEATURES

Location/Qualifiers

1..462  
/Organism="Homo sapiens"

STS  
primer\_bind 49..170

BASE COUNT 105 a 85 c 95 g 141 t 36 others

## ORIGIN

Query Match 6.2%; Score 18; DB 14; Length 462;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TCTGAATTGCTGCTG 37

|||||

Db 138 TCTGAATTGCTGCTG 121

## RESULT 11

AC006427/c 190000 bp DNA HTG 08-JUN-1999

LOCUS Homo sapiens chromosome 4, WORKING DRAFT SEQUENCE, 12 unordered

DEFINITION

pieces.

AC006427

NTD 95019259

VERSION AC006427.4 GI:5019269

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 190000)

Stone, N.E., Schmutz, J.J., Cox, D.R. and Myers, R.M.

Direct Submission

Unpublished

2 (bases 1 to 190000)

Stone, N.E., Schmutz, J.J., Cox, D.R. and Myers, R.M.

Direct Submission

Submitted (21-JAN-1999) Department of Genetics, Stanford Human

Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA

On Jun 8, 1999 this sequence version replaced gi:4966377.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1122: contig of 1122 bp in length

1123 2793: gap of unknown length

2794 3969: contig of 1176 bp in length

3970 5639: gap of unknown length

5640 7494: contig of 1855 bp in length

7495 9165: gap of unknown length

9165 11251: contig of 2087 bp in length

11252 12921: gap of unknown length

12922 24188: contig of 11267 bp in length

24189 25858: gap of unknown length

25859 33031: contig of 7173 bp in length

33032 34701: gap of unknown length

34702 46041: contig of 11340 bp in length

46042 47711: gap of unknown length

47712 67667: contig of 19956 bp in length

67668 69337: gap of unknown length

69338 89177: contig of 19840 bp in length

89178 90847: gap of unknown length

90848 118409: contig of 27562 bp in length

118410 120079: gap of unknown length

120080 148176: contig of 28097 bp in length

148177 149847: gap of unknown length

149847 190000: contig of 40154 bp in length.

## FEATURES

Location/Qualifiers

1..190000  
/Organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"

BASE COUNT 52922 a 35229 c 34931 g 48469 t 18379 others

## ORIGIN

Query Match 6.2%; Score 18; DB 35; Length 190000;

Best Local Similarity 100.0%; Pred. No. 8.8;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 TCATCAGGAATGAGACT 250

|||||

Db 178254 TCATCAGGAATGAGACT 178237

## RESULT 12

AC006034/c 56479 bp DNA PRI 14-MAY-1999

LOCUS Homo sapiens clone NH0144M13, complete sequence.

DEFINITION

AC006034

NTD 94827318

VERSION AC006034.2 GI:4827318

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 56479)

Waterston, R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 56479)

Waterston, R.H.

Direct Submission

Submitted (22-NOV-1998) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

3 (bases 1 to 56479)

Waterston, R.H.

Direct Submission



gene 3779. .5236 /gene="yfmr"  
CDS 3779. .5236 /function="unknown"  
/note="similar to benzaldehyde dehydrogenase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12564.1"  
/db\_xref="PID:el182725"  
/db\_xref="GI:2633059"  
/db\_xref="SPTREMBL:O06478"  
/translation="MEQYELNKPQIGGKMGSSPNVLEKNPPTQKFTTFPRATA  
DWDVATRAALAKKMDVNPFEKRTILEKAVTIEENEAITLMELEGTRKA  
AFELVKNRIKEATFPIRMGKILPSTIDENKRLYRVPAGVGVISPFNPEFLS  
MKSVALGAGNGVVLKPHETPICGTLIAIFENAGIPACILNVVYDIAEIGDSE  
VEHPVRIISFTSGTSGVSIQOLAMKHKRKLLELGGNSAFIVLEDAIEAVNAAY  
FSRFTHOGICMSANRVLVHSSIDYKLELYQAKVESLKVGDPMDDTIGPLINSRO  
TDGLMKYDQAEAGAVPYKLGCGNCTIYEPILIDYKRFMSIAEELGFPVSPKPK  
DSEDAVIDIANETPPGLSGAVTSNLSERGAFAKNTETGMINHTTINDEPNVAFGG  
EKQSGRLNGMSLEETFTLKWISVQHKRSFPY"  
gene 5250. .6136 /gene="yfmS"  
CDS 5250. .6110 /gene="yfmS"  
/function="unknown"  
/note="similar to methyl-accepting chemotaxis protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12565.1"  
/db\_xref="PID:el182726"  
/db\_xref="GI:2633060"  
/db\_xref="SPTREMBL:O06477"  
/translation="MELTINPERKETADIDAFIKAPYLSLVODDITIGIYDEKEL  
VNIAPKTSFLNAGDPLQEDDITDAISNOKTSMVREKELGFLIRARIPLEDEN  
GRVIGVGIGTLESSKLDHYAESLSAVROTAALISDSINGFSWOMGTSIOA  
KRVESAGETIADISTYKGISDQSNILGNAIEARAESKGSVADERTKLATH  
SKENVGQIDQITRKILHSLKLEESIESINHTDQAAVEQISATMOEISSAOHLA  
KMAEKALBEE"  
terminator 6113. .6136 /gene="yfmS"  
gene 6245. .8159 /gene="yfmR"  
CDS 6245. .8134 /gene="yfmR"  
/function="unknown"  
/note="similar to ABC transporter (ATP-binding protein)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB12566.1"  
/db\_xref="PID:el182727"  
/db\_xref="PID:q2633061"  
/db\_xref="GI:2633061"  
/db\_xref="SPTREMBL:O06476"  
/translation="WSILKAENLYTYGDKTLFDHISFHEENRIGLIGPNGTKST  
LKVLAGESEIEGEITKSGSVQVEFLDHPPLPAGQVLEHISGESVMKLTREYE  
KALYELGDPENEGROKHLAAQAKMDANNAMDANTLAKTVLSKGVNDVTPVNEIS  
GGOKRVAIAKLIOPADLLIDEPNHLNDETIEMLGELYSGVAGVALVYDHFVLS  
NRTNRIEYELERGSLEYTKGNEYFLERAEQAFOREKTRONLIRRELALRGA  
KARSTOKARIDRVETLEKQTPQSSGSDPRFISGRHKGQVIEANVNIATDGRMLY  
DRFNLVIGERIGITIGPNGIKTTLNLAGRHHPDGDITIGQVRIQVITQDHS  
ANGELKVIDYIKETAEVVKTAGDMITAEQMLERLFPFSMQOTIRIKLSGEEKRLY  
LLQVLEQPNVFLDEPNTDLDETILSVLEDIDQFPGVITVSHDRFLDVPVRLI  
VLEGNGVLSREOGYSYDVEESAKKAPAPAEKTAEAEPKRRKRLSYDQLEMD  
GIEDKIAQLEKHEQLADIAAGSDPFGIQLMAEQAKTAELEAADRWELESLMI  
EELFS"  
terminator 8138. .8159 /gene="yfmR"  
gene 8257. .8740 /gene="yfmQ"  
CDS 8257. .8703 /gene="yfmQ"

Query Match 5.9% Score 17; DB 1; Length 208430;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 250 TGTCTCTCAGCGCTCCA 266  
DB 99303 TGTCTCTCAGCGCTCCA 99287  
RESULT 14  
D78508/c 9444 bp DNA BCT 05-FEB-1999  
LOCUS D78508.1 GI:1817531  
DEFINITION Bacillus subtilis DNA for yf10, yf1P, yf1N, yf1M, yf1L, yf1K, yf1J,  
yf1I, yf1H, complete cds.  
ACCESSION D78508  
NID 91817531  
VERSION D78508.1 GI:1817531  
KEYWORDS yf1H; yf1I; yf1J; yf1K; yf1M; yf1N; yf1O; yf1P.  
SOURCE Bacillus subtilis (strain:AC327) DNA.  
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
Bacillus.  
REFERENCE 1 (bases 1 to 9444)  
AUTHORS Sekiguchi, J.  
TITLE Direct Submission  
JOURNAL Submitted (27-NOV-1995) to the DDBJ/EMBL/GenBank databases. Junichi  
Sekiguchi, Textile Science and Technology, Shinsu University, 386,  
Department of Applied Biology, 3-15-1 Tokida, Ueda-shi, Nagano 386,  
Japan (E-mail:jseki@uqipc.shinsu-u.ac.jp, Tel:0268-21-5344,  
Fax:0268-21-5331)  
2 (bases 1 to 9444)  
AUTHORS Sekiguchi, J.  
TITLE Nucleotide sequence of Bacillus subtilis chromosome around 77  
degree region  
JOURNAL Unpublished (1996)  
REFERENCE 3 (sites)  
AUTHORS Yamamoto, H., Uchiyama, S. and Sekiguchi, J.  
TITLE The Bacillus subtilis chromosome region near 78 degrees contains  
the genes encoding a new two-component system, three ABC  
transporters and a lipase  
JOURNAL Gene 181 (1-2), 147-151 (1996)  
MEDLINE 97128783  
FEATURES  
source location/Qualifiers  
1. .9444 /organism="Bacillus subtilis"  
/strain="AC327"  
/db\_xref="taxon:1423"  
102. .107  
117. .1058  
/codon\_start=1  
/transl\_table=11  
/product="yf1H"  
/protein\_id="BA011398.1"  
/db\_xref="PID:d1012063"  
/db\_xref="PID:91817532"  
/db\_xref="GI:1817532"  
/translation="WKLSTVDSLGHLPEDMLNFAAKIGIDTLEMTTGGSPAPHLN  
LDLELQSSSEKKEKESQALEKRNMTLCALNCSGNLPDPELQKSHRDITDKTMELAGL  
GKVKYVMSGIPAGGPDQKVPNMITYVSPVPLKDIINQMEDVAIPYKRLYKAE  
ACGVKALLENSSOLVYNPPTLPELRVAVGPMGLNDPESHLLMAGADPIIAARELG  
SAIHVHKDVRIRERYLSAVNGILETRKVTQPARANMYVAVGGQDLOMKKEPFSYV  
KMGYEGEVSLEMDLTPSPAGRTSVALKOTISQ"  
RBS 1055. .1063  
CDS 1068. .2249 /codon\_start=1  
/transl\_table=11  
/product="yf1I"

```
/protein_id="BA011399.1"
/db_xref="PID:d1012064"
/db_xref="GI:1817533"
/translation="MMLNGEIRISKPLRNAMGGGRLSVOYKHKRIGALRNDTAROLT
AGAFDIDAEKRGKDFVNLGYDARCYPNQTFMAEAKRODGLIEVYSIATPNCSTHVEI
KRALAEVHVGICRPLFTSAGEOEIKALAEKKGITGVYTGSGNOMLDMRMATE
OGMTGDIKRVYDIOTROGECADDEGEKISAQKRWVDAIAGSEFVLGDLSTHTYMSQ
LIMPKMIKELCDROSGFSGRAPELNDNAHLYHENGAVGWTMTSSINAGCGDHRI
RIVGSKASIEWMDSKPNELTYEVQEGCOTLVHGMPLDVCNADERLGAHSEGLSE
AMANIYLFKALAIADAKNREDEILNNIYTPIDNAGIEGIMWENCVRSANQSVWVEF
K"
2263..2290
RBS
2360..2368
CDS
2373..3575
/codon_start=1
/transl_table=1
/product="yflJ"
/protein_id="BA011400.1"
/db_xref="PID:d1012065"
/db_xref="GI:1817534"
/translation="MKALFTREMTLWVSCIMLYSIYKEDNMEGYEFLAAGAMITAN
HYLTKRENAIWECLIDIAIGSGEFGIPGTGLFIMLCYAVAFELRGPCKRAMSV
ICLSLITLTVLIRRYAMGNEVDHLSMTFVCGVGLIRKLIDQADQALQOL
OELTSHLALSAHOELHLYAKOVBELTAIERNRMAEIHDTGKMTALLVOLKIL
REMOKRDSOKADEIVGCEETLREKALDVRLSVPTLORENDPSLIESIKOLTEDFCN
AGYTEFFAVSGDPAIIPLSLHPTLLRTVQELTNKRKGGMAACSILOACTTDSISY
IKDDGKNEPEALGFGLNMKRRAAHGGMIRFESERDGFVNAERSLANKKNSFEP
VOOKESLS"
3557..3563
RBS
3572..4234
/codon_start=1
/transl_table=1
/product="yflK"
/protein_id="BA011401.1"
/db_xref="PID:d1012066"
/db_xref="PID:91817535"
/db_xref="GI:1817535"
/translation="MKITITDDODIVRSGLASILQIRELDVITARNQGEAFKAK
ELEPDIYLMDIRMPVNSGVGTGLTSSIPSVYMLLTFKDSALIAELLEGAGYL
LKMDSDITVYKAVTWVSGGWLPELITQMLNEMKREKOLGNGINEPNDLLTE
RELEVIAELGYGNKEIAEKLYITGEIVANKHVSNTISLAVRDRIOALYSVRGVS
VF"
4265..4267
RBS
4362..4368
CDS
4378..5313
/codon_start=1
/transl_table=1
/product="yflL"
/protein_id="BA011402.1"
/db_xref="PID:d1012067"
/db_xref="PID:91817536"
/db_xref="GI:1817536"
/translation="MLQAEINKKAYGKRTIYKISFLKSGSEFGILGPNACKSTTI
SMISGLVPHDSGNITGVGVYKETAAROKIGVPEIYALYPTLNAHEIMWGMXY
GLTHDEAKRAAEVLEYGLTERAKDKITFPGSGMKRINIGALMKRPILLMDPPT
VGIDPDSRNHILEYKQNLNETGMYIITSHMEVEYERLCRIGIIDGGEAIEATGTD
LCSRLGDTIILQTVSGINEAFVAINSLAVNVTYVHELEKIDISAAHHEVYVSL
LAETARHINILISLQVCEPNLERLNLGTGLRD"
5313..5318
RBS
5326..6516
/codon_start=1
/transl_table=1
/product="yflM"
/protein_id="BA011403.1"
/db_xref="PID:d1012068"
/db_xref="PID:91817537"
/db_xref="GI:1817537"
/translation="MKKSTIAMKDVKIRITDRKGFMMILMPLITLCILGALGSVV
K/translation="MKKSTIAMKDVKIRITDRKGFMMILMPLITLCILGALGSVV
DGGSDIDIKVGIOSDOSDITNMFYKDKSIKTYKSGKMKMLIDKKRIDV
GIVIPNHMERKISAVVNAAPDQTLKSSITTEIYAASSFTIOYKAVAEAGSGMDYISKT
EAVKQCKLDPAPFAEKIATLLEKETGDKLITAEKSVGSKATISFOYVSAAMLCKMMLF
```

```
termiator
CDS
complement(7760..8308)
/codon_start=1
/transl_table=1
/product="yflO"
/protein_id="BA011405.1"
/db_xref="PID:d1012070"
/db_xref="PID:91817539"
/db_xref="GI:1817539"
/translation="MRVLYKYLILLLKRGELSGYDITSYKREELGOFMSAKHSQIYPE
IKTLNDEGITEFRITIGTKLEKKNYTLTDSGRKLELDNMLIRHQPPEYKADPEMKA
YFISLQISQEAESDLEFDQLOKROKNSDQSGYTEKLWASAPRMSFSPDGHLVLTG
ALEREKNYVSWLESILAMIDRD"
8569..8573
RBS
8581..9213
/codon_start=1
/transl_table=1
/product="yflP"
/protein_id="BA011406.1"
/db_xref="PID:d1012071"
/db_xref="PID:91817540"
/db_xref="GI:1817540"
/translation="MKKYVMAFIIILSLISVLAAPSGAKAESYHNPVLYVHISGA
SYNFAIKNYLLISQWDSKRLAIIDFYKDTGNNLNGPOLASYDVRVLYKETGAKKVDI
VAHSMGANTLYIKTYLGGCKIKIONVYITAGSNGCYSTALPGTDPNOKIITYTSYSL
NDQIYINSLRLQAGARNIQLYGIGHGLNSQVNGYIKKEGLNGGLMTN"
9224..9258
termiator
BASE COUNT 2654 a 1996 c 2376 g 2418 t
ORIGIN
Query Match 5.9%; Score 17; DB 1; Length 9444;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 250 TGTCTCTCAGCGCCCTCA 266
DB 1179 TGTCTCTCAGCGCCCTCA 1163
RESULT 15
BOVNEURXIA 6349 bp mRNA MM 09-AUG-1995
LOCUS BOVNEURXIA 6349 bp mRNA, complete cds.
DEFINITION Bos taurus neurexin I-alpha mRNA, complete cds.
ACCESSION L14855
NID 9388560
VERSION L14855.1 GI:388560
KEYWORDS neurexin I-alpha.
SOURCE Bos taurus brain.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
```

REFERENCE 1 (bases 1 to 6349)  
 AUTHORS Ulrich, B., Uskaryov, Y. A. and Sudhof, T. C.  
 TITLE Cartography of neuexins: more than 1000 isoforms generated by  
 alternative splicing and expressed in distinct subsets of neurons  
 JOURNAL Neuron 14 (3), 497-507 (1995)  
 MEDLINE 95209856  
 FEATURES  
 SOURCE Location/Qualifiers  
 1. 6349  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /tissue\_type="brain"  
 1293..1382  
 /evidence="experimental"  
 1293..5885  
 /note="amino acid information: bases 1383 .. 5657  
 extracellular domain; amino acid information: bases 5355  
 .. 5567 O-glycosylation domain; amino acid information:  
 bases 5658 .. 5717 transmembrane domain; amino acid  
 information: bases 5718 .. 5882 cytoplasmic domain;  
 precursor"  
 /codon\_start=1  
 /product="neuexin I-alpha"  
 /protein\_id="AA074123.1"  
 /db\_xref="PIR:G388561"  
 /db\_xref="GI:388561"  
 /translation="MGTAALQGGCFLCLSLLLGCVNELSGLEFFPGAEQGWTRFP  
 KMACCESEMSFOLKRSARGLVLYFDGCDPELLITRGRLQLSFISICAPAT  
 LITDTPVNDGAAHNRIRORFNTLFTFOVAKVYVSKRDMTFSGLVGGVLP  
 ELRAALIKTLASVREPRFKMIRDIRNSSIALPVSGEYKLDDEPNSSGSSPC  
 AGESEGGYCLNGVSVYDDVAVCDSCSTRTGRCDCSDENNREGALAMKDGKS  
 KEDNNEGLAHLMGDDQKEEYIAFKSEYFCYDLSQPISSDETLISRTYLRN  
 GLMTSGSADYVNLAKNGAVSLINLSGAFELVPEVNGKFPDNAMHDVAVRNL  
 ROTSGIGHAMVKLHCSVTIYSDGILTTGGYQEDYTMGSDPFYVGGSPETADLP  
 SPVSNFMGCLKEVYKNNNDVLELSRLAKGDPKMKIGVAFKCEVATLDPTE  
 TPESTFSLPKMKAKTKGISPEPTPEPGILFHSGRHOKDAKHPMIVDEPAI  
 EMLDGHVILILDMSGTIRKIRLQKKVNDGEYHYDPQDSSGTSIVNTLTPTTAP  
 GESQILDLDDELITGLPENKAGLVPEVWALNTGIVGICRDLFDGSKDRROM  
 AEVQSTAGVPCPSRETAAPCLSNPCKNNGMRVDCSOTGYLGRSCREAT  
 VLSYDGSMEKIQLPVVMTEADVLSLRFSORAVGILMAITSRSDATLRLELDGR  
 VALLVNDICIRINCNSKSGPETLFAGYNLNDENMHTVVRVGRGSLKLVDDQAMTG  
 OMAGDHTLEFENITETITERRVYLSVSPNTIGHLSITENGMAYIDLCRKGDDYC  
 ELNARFGRNITADPYTRTKSSYVALATQATYSHMLFFQRTKTSIDBLIYNSGDS  
 NDFIVELVKGLIHTVFDGNGANLTKGSNPKLNDONHNMISDTSNLTIVADIT  
 KITQITAGARNLIDKSDLYIGVAKETYSKPKVHAKEGQGLASVLDGRPLD  
 ISDALFCNGAIEGCEGSPSTGCEGSCNMGVCLQOMGISCDSMTSFGICNDPG  
 TTYIRSGGGQITTKMPMDRSTRADRLAIGFSPYOKRAVAVRVDSSGLDYEHL  
 IHQKIGVKNFGYTDIAEESNAILNGKIVHVFRPSGNAITLOVDSMPVIERPA  
 GNDNERLAIARQRIPIRLGRVYDEMILDKGQULITNSQATITITGKQDQPFQGL  
 SGLYNGKLKLNMAENDANIIVGNVRLVGEVPSMTSTETAMQSEMTSIMEYT  
 TLTATSTARGRKPTPEVSTQTDILVASACPDDEIDICEPSSGGLAPTRAG  
 REPYGSAEVIRESSSTIGMVGIVAAALCILILLYAYKRYRNDGEGSYHDESRNY  
 ISNSAOSGAVYKQPSAKANKKKKREYV"  
 1383..5882  
 /function="presynaptic cell surface protein"  
 /evidence="experimental"  
 /product="neuexin I-alpha"  
 1452..2699  
 /note="major repeat A; putative"  
 /rpt\_family="neuexin repeat"  
 1452..1865  
 /note="left arm Aa of the major repeat A; putative"  
 /rpt\_family="neuexin repeat"  
 1941..2057  
 /note="EGF repeat 1; epidermal growth factor; putative"  
 /rpt\_family="EGF"  
 2065..2172  
 /note="spliceable sequence, splice site 1"  
 /evidence="experimental"  
 2262..2699  
 /note="right arm Ab of the major repeat A; putative"  
 /rpt\_family="neuexin repeat"  
 2475..2519  
 misc\_feature

repeat\_region /note="spliceable sequence, splice site 2"  
 /evidence="experimental"  
 2871..4019  
 /note="major repeat B; putative"  
 /rpt\_family="neuexin repeat"  
 repeat\_unit 2871..3314  
 /note="left arm Ba of the major repeat B; putative"  
 /rpt\_family="neuexin repeat"  
 repeat\_unit 3375..3497  
 /note="EGF repeat 2; epidermal growth factor; putative"  
 /rpt\_family="EGF"  
 repeat\_unit 3513..4019  
 /note="right arm Bb of the major repeat B; putative"  
 /rpt\_family="neuexin repeat"  
 misc\_feature 3709..3735  
 /note="spliceable sequence, splice site 3"  
 /evidence="experimental"  
 repeat\_region 4146..5225  
 /note="major repeat C; putative"  
 /rpt\_family="neuexin repeat"  
 repeat\_unit 4146..4541  
 /note="left arm Ca of the major repeat C; putative"  
 /rpt\_family="neuexin repeat"  
 repeat\_unit 4599..4718  
 /note="EGF repeat 3; epidermal growth factor; putative"  
 /rpt\_family="EGF"  
 repeat\_unit 4815..5225  
 /note="right arm Cb of the major repeat C; putative"  
 /rpt\_family="neuexin repeat"  
 misc\_feature 5080..5169  
 /note="spliceable sequence, splice site 4"  
 /evidence="experimental"  
 misc\_feature 5569..5577  
 /note="spliceable sequence, splice site 5"  
 /evidence="experimental"  
 variation replace(5690,"")  
 /note="a" in variation"  
 BASE COUNT 1680 a 1598 c 1660 g 1411 t  
 ORIGIN

Query Match 5 9%: Score 17; DB 3; Length 6349;  
 Best local Similarity 100.0%; Pred. No. 37;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0;  
 Gaps 0;  
 QY 241 AATGAGCTGCTCTC 257  
 |||||  
 DB 2348 AATGAGCTGCTCTC 2364

Search completed: August 6, 1999, 11:36:19  
 Job time: 6265 sec

This Page Blank (uspto)

---

```

* 35369 35388: gap of unknown length
* 62358 62358: contig of 26970 bp in length
* 62359 62378: gap of unknown length
* 87529 87529: contig of 25151 bp in length
* 87530 87549: gap of unknown length
* 87550 106009: contig of 18460 bp in length
* 106010 106029: gap of unknown length
* 106030 118000: contig of 11971 bp in length
* 118001 118020: gap of unknown length
* 118021 126008: contig of 7988 bp in length
* 126009 126028: gap of unknown length
* 126029 132535: contig of 6507 bp in length
* 132536 132555: gap of unknown length
* 132556 136051: contig of 3496 bp in length
* 136052 136071: gap of unknown length
* 136072 139258: contig of 3187 bp in length
* 139259 139278: gap of unknown length
* 139279 141853: contig of 2575 bp in length
* 141854 141873: gap of unknown length
* 141874 143897: contig of 2124 bp in length
* 143898 144017: gap of unknown length
* 144018 145913: contig of 1896 bp in length
* 145914 145933: gap of unknown length
* 145934 147688: contig of 1755 bp in length
* 147689 147708: gap of unknown length
* 147709 149311: contig of 1603 bp in length
* 149312 149331: gap of unknown length
* 149332 150810: contig of 1479 bp in length
* 150811 150830: gap of unknown length
* 150831 152267: contig of 1437 bp in length
* 152268 152287: gap of unknown length
* 152288 153714: contig of 1427 bp in length
* 153715 153734: gap of unknown length
* 153735 155159: contig of 1425 bp in length
* 155160 155179: gap of unknown length
* 155180 156600: contig of 1421 bp in length
* 156601 158021: gap of unknown length
* 158022 158041: contig of 1401 bp in length
* 158042 158060: gap of unknown length
* 158061 159428: contig of 1387 bp in length
* 159429 159448: gap of unknown length
* 159449 160775: contig of 1327 bp in length
* 160776 160795: gap of unknown length
* 160796 162047: contig of 1252 bp in length
* 162048 162067: gap of unknown length
* 162068 163236: contig of 1169 bp in length
* 163237 163256: gap of unknown length
* 163257 164413: contig of 1157 bp in length
* 164414 164433: gap of unknown length
* 164434 165517: contig of 1084 bp in length
* 165518 165537: gap of unknown length
* 165538 166466: contig of 929 bp in length
* 166467 166486: gap of unknown length
* 166487 167409: contig of 923 bp in length
* 167410 167429: gap of unknown length
* 167430 168301: contig of 872 bp in length
* 168302 168321: gap of unknown length
* 168322 169185: contig of 864 bp in length
* 169186 169205: gap of unknown length
* 169206 170049: contig of 844 bp in length
* 170050 170069: gap of unknown length
* 170070 170900: contig of 831 bp in length
* 170901 170920: gap of unknown length
* 170921 171745: contig of 825 bp in length
* 171746 171765: gap of unknown length
* 171766 172533: contig of 768 bp in length
* 172534 172553: gap of unknown length
* 172554 173309: contig of 756 bp in length
* 173310 173329: gap of unknown length
* 173330 174080: contig of 751 bp in length
* 174081 174100: gap of unknown length
* 174101 174850: contig of 750 bp in length
* 174851 174870: gap of unknown length

```

```

* 174871 175619: contig of 749 bp in length
* 175620 175639: gap of unknown length
* 175640 176387: contig of 748 bp in length
* 176388 176407: gap of unknown length
* 176408 177152: contig of 745 bp in length
* 177153 177172: gap of unknown length
* 177173 177910: contig of 738 bp in length
* 177911 177930: gap of unknown length
* 177931 178667: contig of 737 bp in length
* 178668 178687: gap of unknown length
* 178688 179422: contig of 735 bp in length
* 179423 179442: gap of unknown length
* 179443 180176: contig of 734 bp in length
* 180177 180196: gap of unknown length
* 180197 180925: contig of 729 bp in length
* 180926 180945: gap of unknown length
* 180946 181692: contig of 727 bp in length
* 181693 181692: gap of unknown length
* 181693 182416: contig of 724 bp in length
* 182417 182436: gap of unknown length
* 182437 183159: contig of 723 bp in length
* 183160 183179: gap of unknown length
* 183180 183892: contig of 713 bp in length
* 183893 183912: gap of unknown length
* 183913 184611: contig of 699 bp in length
* 184612 184631: gap of unknown length
* 184632 185327: contig of 696 bp in length
* 185328 185347: gap of unknown length
* 185348 186034: contig of 687 bp in length
* 186035 186054: gap of unknown length
* 186055 186725: contig of 671 bp in length
* 186726 186746: gap of unknown length
* 186746 186820: contig of 75 bp in length.

```

```

FEATURES
source
1. 186820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPC14-55A22"
/chromosome="Xq28"

```

```

BASE COUNT      56794 a 36678 c 36106 g 56119 t 1123 others
ORIGIN

```

```

Query Match      8.2%: Score 18; DB 34; Length 186820;
Best Local Similarity 100.0%: Pred. No. 6.3;
Matches 18; Conservative 0; Mismatches 0; Gaps 0;
OY      54 CAAATCAAAATGCAAT 71
|||||
Db 172717 CAAATCAAAATGCAAT 172700

```

```

Search completed: August 6, 1999, 11:34:36
Job time: 6162 sec

```

```

/product="unknown"
/protein_id="AAC33941.1"
/db_xref="db_xref:q3513724"
/db_xref="GI:3513724"
/translation="MMWTRVTKOKAKAMSGVNTGKGIWSPFRIRKTYPGSGDKMHSIQ
ERRVIRNIGVVDNRTORLNEGFARVETSGGSGSKTSGSKERKPPKQYNNGRRL
GSXDRVRECELTGTRKLRAGVGRSLKTRGREGTGNKSHPSRACESE"
misc_feature
3330..3335
/Note="CAP box"
complement(3316..3321)
TATA_signal
3400..3405
/bound_molecy="TFIID-E11a"
/evidence=not_experimental
3553..3560
/misc_feature
/Note="Inr element; similar to a TATA box"
BASE COUNT 922 a 983 c 897 g 1071 t 6 others
ORIGIN
Query Match 8.2%; Score 18; DB 12; Length 3879;
Best Local Similarity 100.0%; Pred. No. 7.2; 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

QY 136 CACAGCCCAATCCCA 153
|||||
Db 2558 CACAGCCCAATCCCA 2575

RESULT 14
AC004676 201175 bp DNA HTG 10-MAY-1998
LOCUS Homo sapiens chromosome X clone bMXD171, WORKING DRAFT SEQUENCE, 1
DEFINITION Ordered pieces.
AC004676
AC004676.1 GI:3126882
VERSION HTG: HTGS_PHASE2.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Mammalia;
Eukaryota; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 201175)
Chen, E., Brownstein, B.H., States, D.J., Schlessinger, D. and
Mazzarella, R.
Direct Submission
Unpublished (1997)
2 (bases 1 to 201175)
Brownstein, B.H., States, D.J. and Mazzarella, R.
Direct Submission
Submitted (10-MAY-1998) Center for Genetics in Medicine, Box 8232,
Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA
Current status of this project is available at:
'http://www.hbc.wustl.edu/cgm/seq_projects.html'
Submitted by:
Elison Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp.,
850 Lincoln Center Drive,
Foster City, CA 94404 USA
e-mail: elison@genseq.apdbio.com
and
Buddy Brownstein,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232
4566 Scott Avenue,
St. Louis, MO 63110, USA
e-mail: Buddy@genetics.wustl.edu
and

```

```

David J. States,
Institute for Biomedical Computing
Washington University in St. Louis
700 South Euclid Ave.
St. Louis, MO 63108 USA
e-mail: states@hbc.wustl.edu.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved
1 201175: contig of 201175 bp in length.
location/Qualifiers
1..201175
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="bMXD171"
/chromosome="X"
BASE COUNT 60554 a 41437 c 41653 g 57531 t
ORIGIN
Query Match 8.2%; Score 18; DB 34; Length 201175;
Best Local Similarity 100.0%; Pred. No. 6.3; 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

QY 47 AAACAAGCAAAATCAAAA 64
|||||
Db 192394 AAACAAGCAAAATCAAAA 192377

RESULT 15
AC006400 186820 bp DNA HTG 26-FEB-1999
LOCUS Homo sapiens chromosome Xq28 clone RPC14-554N22, WORKING DRAFT
DEFINITION SEQUENCE, 54 unordered pieces.
AC006400
AC006400.4 GI:4309854
VERSION HTG: HTGS_PHASE1.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Mammalia;
Eukaryota; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 186820)
Muzny, D., Aronson, A.D., Brundage, E., Carvelli, K., Chen, E., Di, W.,
Ding, Y., Dugan, S., Durbin, T., Forcum, J., Ganesh, R., Garcia, C.,
Goodman, M., Gorrell, J.H., Haywood, M., Jackson, L., Kampal, R.,
Karpathy, S., Leal, B., Li, Y., Liu, W., Logan, O., Lu, J., Ly, T.,
Martinez, C., Oswal, G., Perez, L., Rashid, N.D., Rowland, K.,
Savage, L., Scherer, S.S., Shen, H., Timms, K.M., Todd, J., Vo, Q.,
Worley, K.C., Yu, W., Chinault, C., Nelson, D. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 186820)
Worley, K.C.
Direct Submission
Submitted (15-JAN-1999) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 1, 1999 this sequence version replaced gi:4263633.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 35368: contig of 35368 bp in length

```



QY 173 TGGACAAAGATTGTGTG 190  
|||||  
Db 45076 TGGACAAAGATTGTGTG 45093

RESULT 13  
AF078112 3879 bp DNA ROD 03-SEP-1998  
LOCUS AF078112 Mus musculus calcium signal-modulating ligand (Cam1g) gene,  
DEFINITION promoter region and partial cds; and unknown genes.  
ACCESSION AF078112  
MID 935137719  
VERSION AF078112.1 GI:3513719  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3879)  
AUTHORS Morales,V.M., Snapper,S.B. and Blumberg,R.S.  
TITLE Sequence of the promoter region of the mouse CAM1G gene  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3879)  
AUTHORS Morales,V.M., Snapper,S.B. and Blumberg,R.S.  
TITLE Direct Submission  
JOURNAL Submitted (14-JUL-1998) Gastroenterology, Brigham & Women's  
Hospital, 75 Francis Street, Boston, MA 02115, USA  
Location/Qualifiers  
1. 3879  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="13"  
complement(328..660)  
/note="ORF1"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAC33937.1"  
/db\_xref="PID:g3513720"  
/db\_xref="GI:3513720"  
/translation="MAVLEICRPSKLFSSQOQKNTPPPKRSQHKTPSMPCILREFF  
VLSERGVTCQGLRQNLTOQLMLAMDSLCTRLASNSGASSCCAECLVCVEARRGSOI  
TRNGCYRO"  
TATA\_signal complement(936..841)  
CDS complement(1084..1245)  
/note="ORF2"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAC33938.1"  
/db\_xref="PID:g3513721"  
/db\_xref="GI:3513721"  
/translation="MGMSYXLYKHLNGOCCLSHFSIAVRRHHDGNSYKRLLLRAC  
LEFWRVISP"  
TATA\_signal complement(1139..1143)  
misc\_feature 1254..1261  
repeat\_region /note="1rr element; similar to a TATA box"  
1347..1484  
/rpt\_type=tandem  
/rpt\_unit=CT  
1348..1357  
protein\_bind /bound\_moiety="Pyr factor"  
protein\_bind /bound\_moiety="not\_experimental"  
2055..2060  
protein\_bind /bound\_moiety="C-Myb"  
misc\_signal /evidence="not\_experimental"  
2077..2084  
/note="1rr element; similar to a TATA box"  
2092..2097  
protein\_bind /bound\_moiety="H3"  
TATA\_signal /evidence="not\_experimental"  
2161..2165  
repeat\_region 2173..2328  
/note="81 repeat"

repeat\_region complement(2179..2320)  
CDS /note="H2-MD2 repeat"  
complement(2286..2726)  
/note="ORF3"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAC33939.1"  
/db\_xref="PID:g3513722"  
/db\_xref="GI:3513722"  
/translation="MSEMTVRSRIIFGARRONGCGSGTELLPANASILLGVINGAT  
LSGVALFGGCGDPSLLKSGQOTKPOSTIOSAVHFSMELOKELSLTLTNGVT  
LRTLATPNSIASVARELTIGDNGRGCLCILLVCCFCFRGSRSV"  
2423..2428  
/bound\_moiety="GCN4-ILV1.1"  
/evidence="not\_experimental"  
2446..2454  
/note="GAGA motif"  
2563..2568  
/bound\_moiety="H3"  
/evidence="not\_experimental"  
2578..2584  
/bound\_moiety="Rb (retinoblastoma protein)"  
/evidence="not\_experimental"  
2676..2681  
/bound\_moiety="PEA3"  
/evidence="not\_experimental"  
2758..2764  
/bound\_moiety="E2A"  
/evidence="not\_experimental"  
2778..2783  
/bound\_moiety="F-AC11"  
/evidence="not\_experimental"  
2810..2815  
/bound\_moiety="PEA3"  
/evidence="not\_experimental"  
2824..2833  
/bound\_moiety="ANP"  
/evidence="not\_experimental"  
2832..2841  
/bound\_moiety="p53"  
/evidence="not\_experimental"  
2903..2909  
/note="Spl-1und site"  
2904..2909  
/note="Spl-1E3.1 site"  
2916..2921  
/note="Spl-CS2 site"  
2921..2928  
/bound\_moiety="b1coida"  
/evidence="not\_experimental"  
2982..2988  
/bound\_moiety="NF-S"  
/evidence="not\_experimental"  
3012..3177  
/gene="Cam1g"  
3012..>3177  
/gene="Cam1g"  
/product="Cam1g"  
3012..>3177  
/gene="Cam1g"  
3012..>3177  
/gene="Cam1g"  
/note="Cam1g"  
/codon\_start=1  
/product="calcium signal-modulating ligand"  
/protein\_id="AAC33940.1"  
/db\_xref="PID:g3513723"  
/db\_xref="GI:3513723"  
/translation="MEPVPAATDGGERPATPSGLSASORRAELRRKLLMSSEORINR  
IMGFHRPSSGS"  
complement(3145..3606)  
/note="ORF4"  
/codon\_start=1

repeat\_region 24337. 24630  
/rpt\_family="AluJo"  
repeat\_region 24884. 25155  
/rpt\_family="AluJo"  
repeat\_region complement(25168. 25269)  
/rpt\_family="AluJo"  
repeat\_region 25370. 25565  
/rpt\_family="AluSx"  
repeat\_region complement(25588. 25643)  
/rpt\_family="AluJo"  
repeat\_region 25734. 26025  
/rpt\_family="AluJo"  
repeat\_region 26026. 26099  
/rpt\_family="AluJo"  
repeat\_region complement(26148. 26258)  
/rpt\_family="FLAM\_A"  
repeat\_region 26263. 26876  
/rpt\_family="L1MB8"  
repeat\_region 26900. 27201  
/rpt\_family="AluJo"  
repeat\_region 27504. 27807  
/rpt\_family="AluSg"  
repeat\_region complement(28258. 28557)  
/rpt\_family="AluJo"  
repeat\_region complement(28585. 28717)  
/rpt\_family="FLAM\_C"  
repeat\_region 28817. 29421  
/rpt\_family="LINE2"  
repeat\_region complement(29422. 29584)  
/rpt\_family="AluJo"  
repeat\_region complement(29585. 29872)  
/rpt\_family="AluSg1"  
repeat\_region complement(29885. 30000)  
/rpt\_family="FLAM\_C"  
repeat\_region complement(30021. 30378)  
/rpt\_family="MRT1D"  
repeat\_region 30382. 30672  
/rpt\_family="AluJo"  
repeat\_region complement(30676. 30771)  
/rpt\_family="MER21B"  
repeat\_region complement(30766. 30926)  
/rpt\_family="MER21B"  
repeat\_region complement(30932. 31225)  
/rpt\_family="AluSg"  
repeat\_region complement(31238. 31491)  
/rpt\_family="MER21B"  
repeat\_region 31536. 31840  
/rpt\_family="AluJo"  
repeat\_region complement(31844. 32116)  
/rpt\_family="MER21B"  
repeat\_region 32289. 33205  
/rpt\_family="MER11A"  
repeat\_region 32586. 33205  
/rpt\_family="MER11B"  
repeat\_region 33333. 33440  
/rpt\_family="MTR"  
repeat\_region complement(34103. 34215)  
/rpt\_family="AluJo"  
repeat\_region complement(34223. 34523)  
/rpt\_family="AluJo"  
repeat\_region 34562. 34659  
/rpt\_family="L1MC4"  
repeat\_region 34850. 34931  
/rpt\_family="TAAA)n"  
repeat\_region 34934. 35228  
/rpt\_family="AluJo"

Query Match 8.2% Score 18; DB 11; Length 40247;  
Best Local Similarity 100.0%; Pred. No. 6.7; 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0;  
QY 20 AGAACAAACCAACAA 37  
|||||

DB 15507 AGAACAAACCAACAA 15524  
RESULT 12  
AC005817 188708 bp DNA ROD 06-MAY-1999  
LOCUS , complete sequence.  
DEFINITION AC005817  
AC005817  
VERSION 94755205  
AC005817.5 GI:4755205  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Chen, F., Do, T., Reeves, R.H. and Roe, B.A.  
TITLE 1 (bases 1 to 188708)  
JOURNAL Mus musculus Chromosome 16 BAC Clone b581111 In DCCR Region  
2 (bases 1 to 188708)  
AUTHORS Chen, F., Do, T., Reeves, R.H. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (14-OCT-1998) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
3 (bases 1 to 188708)  
AUTHORS Chen, F., Do, T., Reeves, R.H. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (09-DEC-1998) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
4 (bases 1 to 188708)  
AUTHORS Chen, F., Do, T., Reeves, R.H. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (15-DEC-1998) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
5 (bases 1 to 188708)  
AUTHORS Chen, F., Do, T., Reeves, R.H. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (21-APR-1999) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
6 (bases 1 to 188708)  
AUTHORS Chen, F., Do, T., Reeves, R.H. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (03-MAY-1999) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
7 (bases 1 to 188708)  
AUTHORS Chen, F., Do, T., Reeves, R.H. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAY-1999) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
8 (bases 1 to 188708)  
AUTHORS Chen, F., Do, T., Reeves, R.H. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (06-MAY-1999) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
COMMENT On May 6, 1999 this sequence version replaced gi:4754018.  
FEATURES  
SOURCE location/Qualifiers  
1. 188708  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
BASE COUNT 48957 a 44777 c 46064 g 48910 t  
ORIGIN  
Query Match 8.2% Score 18; DB 12; Length 188708;  
Best Local Similarity 100.0%; Pred. No. 6.3; 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0;

Query Match 8.28; Score 18; DB 11; Length 155382;  
 Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Or 48 AACAGCAAAATCAAAA 65  
 ||||||||||||||||  
 Db 124015 AACAGCAAAATCAAAA 124032

RESULT 11  
 LOCUS HSCOS10 40247 bp DNA PRI 06-DEC-1998  
 DEFINITION Homo sapiens chromosome 3 subtelomeric region.  
 ACCESSION AF109718  
 MID 93978547  
 VERSION AF109718.1 GI:3978547  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eukaryota; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 40247)  
 AUTHORS Trask,B.J., Rowen,L., Nguyen,O., Massa,H.F., Giorgi,D.,  
 Brand-Arpon,V. and Rouquier,S.  
 TITLE Sequence of a genomic duplication found at the subtelomeric region  
 of several chromosomes  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 40247)  
 AUTHORS Rowen,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-NOV-1998) Department of Molecular Biotechnology, Box  
 357730 University of Washington, Seattle, Washington 98195, USA  
 COMMENT Sequencing methodology: High redundancy shotgun. Interspersed  
 Repeats were identified with RepeatMasker (available from  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html).  
 FEATURES  
 Location/Qualifiers  
 1..40247  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /clone="6p21"  
 /clone\_11b="L103NC01"  
 674..816  
 /rpt\_family="L1A3\_Send"  
 876..1076  
 /rpt\_family="L1MA5"  
 1073..1475  
 /rpt\_family="L191\_Send"  
 1078..1836  
 /rpt\_family="L143\_Send"  
 complement(1974..2002)  
 /rpt\_family="AT\_rich"  
 complement(2390..2499)  
 /rpt\_family="LINE2"  
 complement(2624..2893)  
 /rpt\_family="AlusX"  
 complement(3330..3426)  
 /rpt\_family="LINE2"  
 complement(3438..3735)  
 /rpt\_family="Alusq"  
 complement(4234..4517)  
 /rpt\_family="AlusX"  
 complement(4520..4630)  
 /rpt\_family="FLAM\_C"  
 4631..4733  
 /rpt\_family="FLAM\_C"  
 4841..5146  
 /rpt\_family="FLAM\_C"  
 5388..5590  
 /rpt\_family="Alusq"  
 5598..5742  
 /rpt\_family="L191\_Send"  
 /rpt\_family="AlusX"

repeat\_region 5745..5818  
 /rpt\_family="Alusq"  
 complement(6298..6547)  
 /note="968 nt similarity to chimp telomeric DNA clone  
 pcht9 in X74283"  
 6389..8826  
 /note="similarity to human mRNA for KIAA0306 gene in  
 AB002304"  
 7862..8159  
 /note="988 nt similarity to H.sapiens telomeric DNA clone  
 160RELO22 in Z96324"  
 complement(8921..9220)  
 /rpt\_family="Alusq"  
 9789..10238  
 /note="match: 3' ESTs A1200936, A1015622, A1223153  
 Soares-testis cDNA"  
 11047..13152  
 /note="48-base repeat region"  
 complement(14390..14809)  
 /rpt\_family="MERIA"  
 15232..15515  
 /rpt\_family="Alub"  
 complement(15764..15872)  
 /rpt\_family="Alub"  
 15880..16108  
 /note="similarity to H.sapiens telomeric DNA sequence  
 7PRELO27 in Z96680"  
 complement(16109..16408)  
 /rpt\_family="Alub"  
 16520..16592  
 /rpt\_family="MER2"  
 16622..16808  
 /rpt\_family="Alub"  
 17415..17549  
 /rpt\_family="Alub"  
 complement(17804..18223)  
 /rpt\_family="MERIA"  
 18306..18591  
 /rpt\_family="FLAM\_A"  
 18591..18662  
 /rpt\_family="Alub"  
 18672..18880  
 /rpt\_family="Alub"  
 complement(18974..19359)  
 /rpt\_family="MER2B"  
 complement(19385..19693)  
 /rpt\_family="Alub"  
 19764..20066  
 /rpt\_family="Alub"  
 complement(20068..20310)  
 /rpt\_family="MER2B"  
 20475..20586  
 /rpt\_family="MIR"  
 complement(21267..21368)  
 /rpt\_family="Alub"  
 21402..21531  
 /rpt\_family="Alub"  
 complement(21540..21971)  
 /rpt\_family="MERIA"  
 22331..22621  
 /rpt\_family="Alub"  
 22624..22698  
 /rpt\_family="TA)n"  
 22714..23008  
 /rpt\_family="Alusq"  
 23095..23401  
 /rpt\_family="Alub"  
 23423..23574  
 /rpt\_family="LINE2"  
 23625..23985  
 /rpt\_family="LINE2"  
 24177..24330  
 /rpt\_family="MT1D"

mailto:egreenhghri.nih.gov , or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. For further details, see <http://bacpac.med.buffalo.edu/>. The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).

## VECTOR: pCYPAC2

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is D10981816: the clone sequenced to the right is D1159C10. Actual start of this clone is at base position 1 of D10870F17; actual end is at 155382 of D10870F17.

## FEATURES

```
source
  location/Qualifiers
  1..155382
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /map="7q33-q36"
    /clone_1lb="RPCI-5"
    /clone="D10870F17"
    1..361
      /rpt_family="L1"
      803..1029
        /rpt_family="Alu"
        2035..2204
          /rpt_family="L1"
          2210..2377
            /rpt_family="L1"
            2378..2791
              /rpt_family="MALR"
              2792..2961
                /rpt_family="L1"
                3050..3465
                  /rpt_family="L1"
                  3471..3858
                    /rpt_family="L1"
                    3932..4937
                      /rpt_family="L1"
                      5732..6026
                        /rpt_family="Alu"
                        7143..7449
                          /rpt_family="Alu"
                          7705..7939
                            /rpt_family="L2"
                            8442..8513
                              /rpt_family="MER2_type"
                              8550..8605
                                /rpt_family="MER2_type"
                                8618..9335
                                  /rpt_family="MER2L_g"
                                  9384..9502
                                    /rpt_family="MALR"
                                    9869..9931
                                      /rpt_family="Retroviral"
                                      9979..10035
                                        /rpt_family="MALR"
                                        10286..10592
                                          /rpt_family="Alu"
                                          13148..13222
                                            /note="match to EST AA140531 (NID:g1700092) mq03g11.r1"
                                            complement(13150..13210)
                                              /note="match to EST A1020949 (NID:g3235285) ub06g06.r1"
                                              complement(13155..13212)
                                                /note="match to EST A1020949 (NID:g3235285) ub06g06.r1"
                                                13156..13201
                                                  /note="match to EST AA261046 (NID:g1897566) mz81a02.r1"
                                                  complement(13157..13212)
                                                    /note="match to EST AA818166 (NID:g2886046)"
                                                    complement(13160..13211)
                                                      /note="match to EST AA818166 (NID:g2886046)"
```

```
misc_feature complement(13164..13210)
  /note="match to EST AA128837 (NID:g1688718) z010d09.r1"
misc_feature 13166..13209
  /note="match to EST AA684287 (NID:g2670873) vm54b03.r1"
misc_feature 13167..13210
  /note="match to EST AA261046 (NID:g1897566) mz81a02.r1"
misc_feature 13167..13212
  /note="match to EST AA140531 (NID:g1700092) mq03g11.r1"
misc_feature 13167..13212
  /note="match to EST AA684287 (NID:g2670873) vm54b03.r1"
misc_feature 13169..13209
  /note="match to EST AA247116 (NID:g1878557)"
misc_feature complement(13829..13845)
  /note="similar to EST AA134277 (NID:g1691756) z006a09.r1"
repeat_region 13841..13987
  /rpt_family="L1"
repeat_region 14055..14342
  /rpt_family="L1"
repeat_region 14836..15084
  /rpt_family="L1"
repeat_region 15262..15562
  /rpt_family="Alu"
repeat_region 15629..16383
  /rpt_family="L1"
repeat_region 16400..16922
  /rpt_family="L1"
repeat_region 16923..17088
  /rpt_family="L1"
repeat_region 17089..17442
  /rpt_family="L1"
repeat_region 17528..17954
  /rpt_family="MALR"
repeat_region 17992..18538
  /rpt_family="L1"
repeat_region 18539..18922
  /rpt_family="Retroviral"
repeat_region 18924..20551
  /rpt_family="Retroviral"
repeat_region 20552..20814
  /rpt_family="Retroviral"
repeat_region 20815..20920
  /rpt_family="L1"
repeat_region 21186..21263
  /rpt_family="L1"
repeat_region 21264..21751
  /rpt_family="Retroviral"
repeat_region 21756..21940
  /rpt_family="L1"
repeat_region 21950..22758
  /rpt_family="L1"
repeat_region 22816..22957
  /rpt_family="L1"
repeat_region 23094..23180
  /rpt_family="MER2_type"
repeat_region 23174..23232
  /rpt_family="MER2_type"
repeat_region 23237..23368
  /rpt_family="MALR"
repeat_region 23369..23562
  /rpt_family="L1"
repeat_region 23563..23774
  /rpt_family="MALR"
repeat_region 23780..23953
  /rpt_family="MER2_type"
repeat_region 23665..25959
  /rpt_family="Alu"
repeat_region 26130..26589
  /rpt_family="L1"
repeat_region 26590..26901
  /rpt_family="L1"
repeat_region 27110..27585
  /rpt_family="MER4-group"
repeat_region 28458..28574
```

```

      intron      229..599
                  /number=1
      exon        600..882
                  /number=2
      intron      883..930
                  /number=2
      exon        931..978
                  /number=3
      intron      979..1025
                  /number=3
      exon        1026..>1111
                  /number=4
BASE COUNT      529 a 285 c 255 g 547 t
ORIGIN
Query Match      8.6%; Score 19; DB 36; Length 1616;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 CAAACAGCAAAATCAAAA 64
      |||
Db 1142 CAAACAGCAAAATCAAAA 1160

RESULT 9
EOY18433 1765 bp DNA INV 20-NOV-1998
LOCUS Euplotes octocarinatus gene encoding pheromone 5, strain EG 20-11.
DEFINITION Y18433
ACCESSION Y3901004
VERSION Y18433.1 GI:3901004
KEYWORDS pheromone 5.
SOURCE Euplotes octocarinatus.
ORGANISM Euplotes octocarinatus.
REFERENCE 1 (bases 1 to 1765)
AUTHORS Moellenbeck, M. and Heckmann, K.
JOURNAL Unpublished
TITLE Direct Submission
AUTHORS Submitted (18-NOV-1998) M. Moellenbeck, Institut fuer Allgemeine
JOURNAL Zoologie und Genetik, Schlossplatz 5, D-48149 Muenster, FRG
FEATURES
SOURCE location/Qualifiers
1..1765
/organism="Euplotes octocarinatus"
/macronuclear
/strain="EG 20-11"
/db_xref="taxon:5937"
join<163..228,743..1028,1079..1126,1172..>1257)
<163..228
/number=1
join(163..228,743..1028,1079..1126,1172..1257)
/codon_start=1
/transl_table=10
/product="pheromone 5"
/protein_id="CAA7174.1"
/db_xref="PID:e1353826"
/db_xref="PID:g3901005"
/db_xref="GI:3901005"
/translation="MKAFITLILALIMVQAQPMKSKVNTKIOSOSKPKOSKRLAST
FMTNLRDAPDCISQYITIGCNTNPDCKWINSNCGSTDSLGECTYNDPENPEVGD
NIAQVIFDRMLGCEYEDVSNCVADGAMVAIFSSOYLNCNGYEFGNVDFYEGCYV
P"
intron 229..742
/number=1
exon 743..1028
/number=2
intron 1029..1078
/number=2
exon 1079..1126
/number=3

```

```

      intron      1127..1171
                  /number=3
      exon        1172..>1257
                  /number=4
BASE COUNT      584 a 296 c 280 g 605 t
ORIGIN
Query Match      8.6%; Score 19; DB 36; Length 1765;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 CAAACAGCAAAATCAAAA 64
      |||
Db 1288 CAAACAGCAAAATCAAAA 1306

RESULT 10
AC004911 155382 bp DNA PRI 28-DEC-1998
LOCUS Homo sapiens PAC clone D0870F17 from 7q33-q36, complete sequence.
DEFINITION AC004911
ACCESSION 93478665
VERSION 93478665
KEYWORDS AC004911.1 GI:3478665
SOURCE HNG.
ORGANISM human.
REFERENCE 1 (bases 1 to 155382)
AUTHORS Bauer, C., Langston, Y., Harrison, M., and Lennox, S.
JOURNAL Unpublished (1998)
TITLE The sequence of Homo sapiens PAC clone D0870F17
AUTHORS Waterston, R.H.
JOURNAL Direct Submission
TITLE Submitted (12-JUN-1998) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 155382)
AUTHORS Waterston, R.
JOURNAL Direct Submission
TITLE Submitted (27-AUG-1998) Department of Genetics, Washington
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 27, 1998 this sequence version replaced gi:3213086.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sepiens@wustl.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7, send

```

repeat\_region 28775..28812 /rpt\_family="(CA)n"  
repeat\_region 28812..28832 /rpt\_family="(TC)n"  
repeat\_region 28813..28963 /rpt\_family="L2"  
repeat\_region complement(28955..29487) /rpt\_family="L2"  
repeat\_region 29488..29798 /rpt\_family="Alusg"  
repeat\_region complement(29799..29961) /rpt\_family="L2"  
repeat\_region 30630..30657 /rpt\_family="(TAAAA)n"  
repeat\_region 31112..31405 /rpt\_family="AluJo"  
repeat\_region 31424..31572 /rpt\_family="AluJo"  
repeat\_region 31573..31874 /rpt\_family="AluJo"  
repeat\_region 31875..32033 /rpt\_family="AluJo"  
repeat\_region 32621..32794 /rpt\_family="MLT1C"  
repeat\_region 32801..33062 /rpt\_family="AluY"  
repeat\_region 33073..33368 /rpt\_family="AlusX"  
repeat\_region 33397..33656 /rpt\_family="MLT1C"  
repeat\_region complement(34645..34822) /rpt\_family="MER91A"  
repeat\_region 35047..35349 /rpt\_family="AluJo"  
repeat\_region 35382..35777 /rpt\_family="AluJo"  
repeat\_region 35970..35990 /rpt\_family="L1MD2"  
repeat\_region 36131..36311 /rpt\_family="NT\_r1ch"  
repeat\_region /rpt\_family="MIR"  
repeat\_region complement(36361..36434)  
Query Match 9.1%; Score 20; DB 42; Length 126732;  
Best Local Similarity 100.0%; Pred. No. 0.46;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 22 AACAAACCAACAAAGAA 41  
DB 9650 AACAAACCAACAAAGAA 9669  
RESULT 7  
AC007160 142178 bp DNA HTG 07-APR-1999  
LOCUS Homo sapiens clone D0146103, WORKING DRAFT SEQUENCE, 2 unordered  
DEFINITION pieces.  
AC007160  
NID 94572709  
VERSION AC007160.2 GI:4572709  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Mammalia;  
Eukaryota; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 142178)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 142178)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

COMMENT MO 63108, USA  
On Apr 7, 1999 this sequence version replaced gi:4508163.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
FEATURES  
source 1 49134: contig of 49134 bp in length  
\* 49135 49152: gap of unknown length  
\* 49153 142178: contig of 93026 bp in length.  
Location/Qualifiers  
1..142178  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="D0146103"  
BASE COUNT 39502 a 33809 c 32673 g 36176 t 18 others  
ORIGIN  
Query Match 8.6%; Score 19; DB 35; Length 142178;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 98 TGATTCGAGGACTTTAG 116  
DB 89023 TGATTCGAGGACTTTAG 89005  
RESULT 8  
EOY17505 1616 bp DNA INV 17-JUN-1998  
LOCUS Euplotes octocarinatus gene encoding pheromone 5.  
DEFINITION Y17505  
ACCESSION 93687407  
NID Y17505.1 GI:3687407  
VERSION 1  
KEYWORDS pheromone 5.  
SOURCE Euplotes octocarinatus.  
ORGANISM Euplotes octocarinatus.  
REFERENCE 1 (bases 1 to 1616)  
AUTHORS Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotes.  
REFERENCE Moellenbeck,M. and Heckmann,K.  
JOURNAL Unpublished  
TITLE 2 (bases 1 to 1616)  
AUTHORS Moellenbeck,M.  
JOURNAL Direct Submission  
Submitted (08-JUN-1998) M. Moellenbeck, Institut fuer Allgemeine  
Zoologie und, Genetik, Schlossplatz 5, D-48149 Muenster, FRG  
Location/Qualifiers  
1..1616  
/organism="Euplotes octocarinatus"  
/macronuclear  
/strain="DM-5"  
/db\_xref="taxon:5937"  
1616  
<163..228  
/number=1  
join(163..228,600..882,931..978,1026..1111)  
/codon\_start=1  
/transl\_table=10  
/codon=(seq: tga,aa:Cys)  
/product="pheromone 5"  
/protein\_id="CAA76771.1"  
/db\_xref="PID:e129594"  
/db\_xref="PID:g3687408"  
/db\_xref="GI:3687408"  
/db\_xref="SPTREMBL:O76943"  
/translation="MKATFTLAIILAVTAFTKTSVNTKIOSKFSKRNKLAST  
FNTNLTAKDAPDYSQYITGNTNPDCKYRNSGCGTNGSDMLYMEKSCVGDN  
IAGYIFDRMAGCYEDVSNVDAAGATLFFSOYLCNCGYEGANNNTYFGCGYIP  
" join(<163..228,600..882,931..978,1026..>1111)  
MRNA

JOURNAL  
REFERENCE  
AUTHORS  
2 Unpublished  
(bases 1 to 126732)  
Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,  
Doneelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,  
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Mejdrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Tesfaye,S., Toriuela-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission  
Submitted (18-MAR-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 126732)  
Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,  
Doneelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,  
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Mejdrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Tesfaye,S., Toriuela-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission  
Submitted (12-APR-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 12, 1999 this sequence version replaced g1:4580489.  
All repeats were identified using RepeatMasker: Smt, A.F.A. &  
Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Only 126732 base pairs from the middle of this clone are being  
submitted. The remainder overlaps either accession number AC003950  
(WICGR project 1261) or accession number AC004584 (WICGR project  
L312).

FEATURES  
source  
1. 126732 Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="hRPK.166\_p.13"  
/clone\_1lb="RPCI-11 human BAC library"  
/map="17"  
/chromosome="17"  
complement(225..399)  
/rpt\_family="MIR"  
634..635  
/note="Unresolved compression. Probably CC."

repeat\_region  
1159..1195  
/rpt\_family="AT\_rich"  
3460..3489  
/rpt\_family="AT\_rich"  
complement(3490..3642)  
/rpt\_family="MER104"  
4211..4233  
/rpt\_family="(CA)n"  
4243..4402  
/rpt\_family="AluJo"  
7083..7483  
/rpt\_family="MER47B"  
7638..7882  
repeat\_region

repeat\_region  
8018..8113  
/rpt\_family="MIR"  
complement(8318..8411)  
/rpt\_family="MER5B"  
8424..8563  
/rpt\_family="MER5A"  
complement(8843..9093)  
/rpt\_family="AluSg"  
9133..9428  
/rpt\_family="AluB8"  
9686..9790  
/rpt\_family="F1AM\_A"  
complement(10090..10230)  
/rpt\_family="L1M4"  
10237..10403  
/rpt\_family="MER5B"  
10410..10677  
/rpt\_family="AluSx"  
10809..10849  
/rpt\_family="MER57B"  
complement(10858..11072)  
/rpt\_family="AluSp"  
11073..11093  
/rpt\_family="AT\_rich"  
complement(11257..11436)  
/rpt\_family="AluSx"  
11437..11475  
/rpt\_family="(TG)n"  
complement(11476..11600)  
/rpt\_family="AluSx"  
12840..13143  
/rpt\_family="AluSg"  
14926..15020  
/rpt\_family="Char11e5"  
16416..16439  
/rpt\_family="(A)n"  
16455..16733  
/rpt\_family="AluB"  
16995..17296  
/rpt\_family="AluB"  
17521..17655  
/rpt\_family="HAL1"  
17794..18087  
/rpt\_family="HAL1"  
18667..18830  
/rpt\_family="MER105"  
complement(18868..19055)  
/rpt\_family="L2"  
19912..20197  
/rpt\_family="AluJo"  
20198..20224  
/rpt\_family="(TAA)n"  
21204..21281  
/rpt\_family="MIR"  
22304..22496  
/rpt\_family="MER3"  
22544..22564  
/rpt\_family="AT\_rich"  
26544..26600  
/rpt\_family="GC\_rich"  
complement(26694..26695)  
/note="Unresolved compression. Probably GC."

repeat\_region  
26955  
/rpt\_family="AluSg"  
28753..28774  
/rpt\_family="L2"  
repeat\_region

```

repeat_region complement(14083, .14322)
repeat_region /rpt_family="MIR"
repeat_region 14484, .14510
repeat_region /rpt_family="AT-rich"
repeat_region complement(15900, .16212)
repeat_region /rpt_family="L1MAA"
repeat_region 16280, .16396
repeat_region /rpt_family="MER7A"
repeat_region 16986, .17306
repeat_region /rpt_family="AluJo"
repeat_region 17422, .17468
repeat_region /rpt_family="(CAAAA)n"
repeat_region 17550, .17642
repeat_region /rpt_family="HAL1"
repeat_region complement(17643, .17937)
repeat_region /rpt_family="AluSp"
repeat_region 17938, .18133
repeat_region /rpt_family="HAL1"
repeat_region 18764, .19102
repeat_region /rpt_family="L1ME"
repeat_region 19121, .19742
repeat_region /rpt_family="L1M4"
repeat_region 20318, .20493
repeat_region /rpt_family="MLT1J"
repeat_region complement(20498, .20831)
repeat_region /rpt_family="L1M4"
repeat_region complement(20888, .21059)
repeat_region /rpt_family="AluSg"
repeat_region complement(21060, .21328)
repeat_region /rpt_family="AluY"
repeat_region complement(21329, .21460)
repeat_region /rpt_family="AluSg"
repeat_region 22123, .23454
repeat_region /rpt_family="L1MB2"
repeat_region 23455, .23586
repeat_region /rpt_family="FLAM_C"
repeat_region 23587, .24009
repeat_region /rpt_family="L1MB2"
repeat_region 24010, .24312
repeat_region /rpt_family="AluSg"
repeat_region 24333, .25000
repeat_region /rpt_family="L1M1"
repeat_region 25001, .25298
repeat_region /rpt_family="AluSg"
repeat_region 25299, .26025
repeat_region /rpt_family="L1M1"
repeat_region 26026, .26321
repeat_region /rpt_family="AluSg"
repeat_region 26335, .27573
repeat_region /rpt_family="L1M1"

```

```

Query Match 9.1%; Score 20; DB 11; Length 158905;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 127 CTAAGACACACAGCCACCA 146
DB 151223 CTAAGACACACAGCCACCA 151242

```

```

RESULT 5
AB017156 AB017156 2937 bp mRNA ROD 04-MAR-1999
LOCUS Mus musculus gob-5 mRNA, complete cds.
DEFINITION AB017156
ACCESSION AB017156
VERSION 93721911
KEYWORDS GOB-5.
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)

```

AUTHORS Komiya,T., Tanigawa,Y. and Hirohashi,S.  
 TITLE Cloning and identification of the gene gob-5, which is expressed in  
 intestinal goblet cells in mice  
 JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
 MEDLINE 99160866  
 2 (bases 1 to 2937)  
 AUTHORS Komiya,T., Tanigawa,Y. and Hirohashi,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tooru  
 Komiya, ERATO, JSR, Hirohashi Cell Configuration Project; 5-9-9,  
 Tokodai, Tsukuba, Ibaraki 300-2635, Japan  
 (E-mail:tkom@hccp.jst.go.jp, Tel:81-298-47-7563,  
 Fax:81-298-47-5226)

FEATURES  
 source Location/Qualifiers  
 1..2937

gene  
 CDS  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_type="goblet cell"  
 /dev\_stage="adult"  
 /tissue\_type="intestine"  
 15..2756  
 /gene="gob-5"  
 15..2756  
 /gene="gob-5"  
 /codon\_start=1  
 /protein\_id="BA33743.1"  
 /db\_xref="PID:d1034712"  
 /db\_xref="PID:g3721912"  
 /db\_xref="GI:3721912"

```

/translacion="MESLKSFPVLLIHLLEGVISELIDLNNGHEGYIAIDHVP
EDEALIOIKDMVQASPYLFQATGKRRFRKNAIILPESWKAAPETRKLEFFKA
DIVLSTSPLENDKEFYTHIGACGCKGIRHLPPDLGAKRLQYGQDRTFHEWMA
FRKVENYNNDEKFTYLSKQCPAASATITGKNOYRQGGSCINGRCVLDVYAG
LYKNCVTPDPHONERKASTMENONINSVYRQVYSLKNSMDLDRNNOASRLF
IQSEDEKQTPHTAOPAPRPSRLQAGRVLYLQKSGADRLIHLTPVSAAGTSGLR
LIQVEQSGWGVTFDSAAVQSELQOLNSGADRLIHLTPVSAAGTSGLR
AFTYIKKKPTDGESEIVLTDGENTLSSCDLVKQGAIIHTYALCPAAKELPQIS
KMTGLQYSSDQVONNGLVDAFALSSGNAIAQHSIQLESQVNLQNNOMNGSVI
VDSVSGKDTLFTWTTHPTPIFIMDSGVQNGFIIIDTTTKVYIAQVPAKGVFWK
ASVTLAISVNGKVTTLLELDNGAGADTKNDGVSYRFTAFDANGRSYIMLGV
TSDRQAPAPKRNKATIDGWIDEDEVARNRPPRPTSYODKQCFESRTSSGGSVATN
VPAAPRIPDPFPCQITDLKASIQGQNLVNLWTFAPGDDYDRGASVYIMPSYSD
LRDHFTSLQVNTGLIKKASSEIFEFELGWTFGNGTDFIATIDAVKSNKSEI
SNIAHVSFIPAOEPIPEDSTPCPDISINSTIIPGIHVLIKMKMWLGEMQVTLGLH"
BASE COUNT 860 a 718 c 693 g 666 t
ORIGIN

```

```

Query Match 9.1%; Score 20; DB 12; Length 2937;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 171 ATTGGACAAAGATTGCTG 190
DB 921 ATTGGACAAAGATTGCTG 940

```

```

RESULT 6
AC007114 AC007114 126732 bp DNA PRI 12-APR-1999
LOCUS Homo sapiens chromosome 17, clone hRPK.166_P.13, complete sequence.
DEFINITION AC007114
ACCESSION AC007114
VERSION 94581193
KEYWORDS HNG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 126732)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone hRPK.166_P.13

```



```

VERSION      AC005666.1  GI:4049331
KEYWORDS     HTG.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 158905)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE        Homo sapiens chromosome 17, clone hRPK.112.H_10
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 158905)
AUTHORS      Birren,B., Fasmann,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
              Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J.,
              Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M.,
              Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E.,
              Devon,K., Dewar,K., Donegan,L., Ferreira,P., Fitzhugh,W.,
              Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraghty,K., Grant,G.,
              Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C.,
              Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J.,
              Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K.,
              Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
              Nafit,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
              Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P.,
              Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
              Subramanian,A., Testfaye,S., Tichonolsky,N., Torruella-Miller,I.,
              Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
              Ye,W.J., Zhao,J. and Zody,M.
TITLE        Direct Submission
JOURNAL      Submitted (10-SEP-1998) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE    3 (bases 1 to 158905)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
              Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Brown,A.,
              Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
              Cooke,P., Dearlano,K., Depayre,E., Devon,K., Dewar,K.,
              Donegan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
              Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A.,
              Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A.,
              Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGuirk,A.,
              McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
              Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
              Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A.,
              Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C.,
              Subramanian,A., Testfaye,S., Torruella-Miller,I., Vassiliev,H.,
              Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J. and Zody,M.
TITLE        Direct Submission
JOURNAL      Submitted (22-DEC-1998) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Dec 22, 1998 this sequence version replaced gi:4028945.
              All repeats were identified using RepeatMasker: Smtt, A.F.A. &
              Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RW/RepeatMasker.html.
FEATURES     Location/Qualifiers
              1..158905
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="RP11 human BAC library"
               /clone_id="RP11-112.H_10"
               /map="17"
               /chromosome="17"
               complement(835..1135)
               /rpt_family="AluSg"
               repeat_region
               complement(1624..1918)
               /rpt_family="AluSg"
               repeat_region
               2302..2467
               /rpt_family="L1MB3"
               repeat_region
               2471..2778
               /rpt_family="AluY"
               repeat_region
               2779..2925
               /rpt_family="AluJc/FRAM"
               repeat_region
               2926..2948
               /rpt_family="(CAA)n"
               repeat_region
               2955..3255

```

```

repeat_region      /rpt_family="L1MB3"
                    3256..3560
repeat_region      /rpt_family="AluSc"
                    3570..3729
repeat_region      /rpt_family="FAM"
                    3730..3787
repeat_region      /rpt_family="L1MB3"
                    complement(3964..4114)
repeat_region      /rpt_family="MER5B"
                    5111..5426
repeat_region      /rpt_family="AluYb"
                    5480..5515
repeat_region      /rpt_family="AT_rich"
                    5752..5784
repeat_region      /rpt_family="AT_rich"
                    complement(5999..6328)
repeat_region      /rpt_family="MER5B"
                    6435..6518
repeat_region      /rpt_family="L1MB5"
                    6519..6815
repeat_region      /rpt_family="AluY"
                    6816..7533
repeat_region      /rpt_family="L1MB5"
                    7535..7694
repeat_region      /rpt_family="AluYb"
                    7705..8001
repeat_region      /rpt_family="AluSg"
                    8124..8183
repeat_region      /rpt_family="MLT2D"
                    complement(8184..8205)
repeat_region      /rpt_family="AluSg"
                    complement(8206..8229)
repeat_region      /rpt_family="(CAA)n"
                    complement(8230..8515)
repeat_region      /rpt_family="AluSg"
                    8516..8942
repeat_region      /rpt_family="MLT2D"
                    complement(8957..9079)
repeat_region      /rpt_family="MER7A"
                    9080..9419
repeat_region      /rpt_family="THE1C"
                    complement(9420..9626)
repeat_region      /rpt_family="MER7A"
                    9702..9748
repeat_region      /rpt_family="MLT2D"
                    9863..9940
repeat_region      /rpt_family="L2"
                    10022..10238
repeat_region      /rpt_family="L1MB3"
                    complement(10325..11142)
repeat_region      /rpt_family="L2"
                    complement(11143..11173)
repeat_region      /rpt_family="(TGA)n"
                    complement(11175..11440)
repeat_region      /rpt_family="AluYb"
                    complement(11441..12055)
repeat_region      /rpt_family="L2"
                    complement(12056..12364)
repeat_region      /rpt_family="AluSg"
                    complement(12365..12447)
repeat_region      /rpt_family="L2"
                    12448..12722
repeat_region      /rpt_family="AluSc"
                    12726..13024
repeat_region      /rpt_family="AluYb"
                    complement(13025..13417)
repeat_region      /rpt_family="L2"
                    13419..13698
repeat_region      /rpt_family="AluX"
                    13699..13726
repeat_region      /rpt_family="(TAAA)n"
                    13783..14076
repeat_region      /rpt_family="AluY"

```

ACCESSION AF039401  
MID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 35278)  
Gruber,A.D., Elble,R.C., Ji,H.L., Schreier,K.D., Fuller,C.M. and  
Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human CIC1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
Gruber,A.D., Elble,R. and Pauli,B.U.  
AUTHORS Direct Submission  
TITLE Submitted (19-DEC-1997) Department of Pathology, College of  
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
LOCATION/Qualifiers  
1. 35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1. 1617  
/gene="hCIC1"  
1. 33522  
/gene="hCIC1"  
767. 1265  
/gene="hCIC1"  
/rpl\_family="L1 transposable element"  
/rpl\_type="dispersed"  
1587. 1596  
/gene="hCIC1"  
/note="putative"  
1618. 1708  
/gene="hCIC1"  
/number=1  
join(1618..1708,1662..2383,6667..6807,6979..7126,  
9696..9801,15454..15631,18591..18809,19783..20010,  
22253..22427,24522..24628,26644..26859,27447..27708,  
28766..28936,31840..32079,32919..33522)  
/gene="hCIC1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCIC1"  
/number=2  
join(2222..2283,6667..6807,6979..7126,9696..9801,  
15454..15631,18591..18809,19783..20010,22253..22427,  
24522..24628,26644..26859,27447..27708,28766..28936,  
31840..32079,32919..33310)  
/gene="hCIC1"  
/protein\_id="AAC35429.1"  
/product="calcium-dependent chloride channel-1"  
/codon\_start=1  
/db\_xref="PIR:G4009460"  
/db\_xref="GI:4009460"

LDKRENSLQVNTALIPREANSEEVLFKPENTFFENGTDLFIAGVADKDKSEI  
SNARVSLFIPQTPPETSPDSTAPCPNHINSTIPGHIILKIMRWIGELQSLIA  
"  
6667..6807  
/gene="hCIC1"  
/number=3  
6979..7126  
/gene="hCIC1"  
/number=4  
9696..9801  
/gene="hCIC1"  
/number=5  
15454..15631  
/gene="hCIC1"  
/number=6  
18591..18809  
/gene="hCIC1"  
/number=7  
19783..20010  
/gene="hCIC1"  
/number=8  
22253..22427  
/gene="hCIC1"  
/number=9  
24522..24628  
/gene="hCIC1"  
/number=10  
26644..26859  
/gene="hCIC1"  
/number=11  
27447..27708  
/gene="hCIC1"  
/number=12  
28766..28936  
/gene="hCIC1"  
/number=13  
31840..32079  
/gene="hCIC1"  
/number=14  
32919..33522  
/gene="hCIC1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 99.1%; Score 218; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 3.9e-114; Indels 0; Gaps 0;  
Matches 218; Conservative 0; Mismatches 0;  
QY 3 ATAGTGAATTCCTGCTAGACAAACCAACCAAGAGCTCCAAACGCAAAATCAA 62  
|||||  
DB 18591 ATAGTGAATTCCTGCTAGACAAACCAACCAAGAGCTCCAAACGCAAAATCAA 18650  
QY 63 AAATGCAATCTCCGAGCAGCATGGAAGTATCCGATTCGAGGACTTTAAGAAAAC 122  
|||||  
DB 18651 AAATGCAATCTCCGAGCAGCATGGAAGTATCCGATTCGAGGACTTTAAGAAAAC 18710  
QY 123 ACTCCTATGACAAACAGCCACCAATCCACCTCTCAATGCTGAGATGGACAAAGA 182  
|||||  
DB 18711 ACTCCTATGACAAACAGCCACCAATCCACCTCTCAATGCTGAGATGGACAAAGA 18770  
QY 183 ATGTGCTTAGCTTGCATCAATCTGGAAGCATGGC 220  
|||||  
DB 18771 ATGTGCTTAGCTTGCATCAATCTGGAAGCATGGC 18808  
RESULT 4  
AC005666 158905 bp DNA PRI 22-DEC-1998  
LOCUS Homo sapiens chromosome 17, clone hRPK.112.H.10, complete sequence.  
DEFINITION AC005666  
ACCESSION 94049331  
NID

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

## FEATURES

source  
1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hclcal"  
352. .3096  
/gene="hclcal"  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="pid:g4009458"  
/db\_xref="gi:4009458"  
/translation="MGPKSSVFLLIHLLEGALSLSLIQLNNNGEIVVAIDPNP  
EDELIIQIKDMYQASLYLEATGKRFYFNVAIIPETWTKADYVRKLEYKNA  
DVLVAESTPPNDPEYTEMNGCKEGERIHLPDFIAGKLAEGPOGKAHEVMAH  
LRMGVFEYNDDEKFLSNGRIQAVRSAGITGVNKKCOGSGCYTRKCTFNKVTGL  
YKCEGFYLSRQTERKASIMFAOHVDSIVFECTEONNKKKAPNOKNCRSMWEV  
RDSDFKRTTPPTTPPTPSLQIGRIYCLVDKSGMATGRNLRNLOAQFL  
LQVLELSWGVAVTFDSAHVQSELIOINSQSDROTAKRLPAASGGSISCSLSRA  
FTVIRKKYPTDGESEIVLLTDEEDNTISGCFNEVKSQSAIHTVALGPSAOLELELSK  
MTGLOIYASPOVONNGILDAFGALSNGAVSORSIOLESGLTLQNSOMNGTVIY  
DSYVKDITFLITWTQPOLILMDPSOGOGFVVDKTKMAVLIQIGIAKGVWXY  
SLQASQTLTLTVTSRASNATIPITYTSKNTKSTKPSPLVYVYANIRGASPIIRA  
SVTALIESVNGKTVLELNDGAGADATKDGYSRTFTYDINGRISVYKRALGVN  
AARRVVIPOQSGALYIPGWIENDEIQNNPPPEINKDVQHRQVCFSTSSGGSFVAS  
DVNPAPIPDLFPPGOITDLKAEIHGSLINLTWAPGDYDGHGAHXYIIRISYILD  
LRDKFNESLQVNTALIPKEANSEVEFLFKNITFENGTDLFIQIADVYDKSEI  
SNIRVSLFIPQPPETPSPDETSAFCPNHINSTIPGHIILKIMMKWIGELQLSIA  
"

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 100.0%; Score 220; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 3,1e-115;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTATAGTTGAATTCGTGTACAGAAACCAACCAAGAGCTCCAAACAGCAAAATC 60  
DB 1085 CTATAGTTGAATTCGTGTACAGAAACCAACCAAGAGCTCCAAACAGCAAAATC 1144  
QY 61 AAAAATGCAATCTCCGACACATGGGAAGTATCGGTATTCGTGAGACTTAAGAAAA 120  
DB 1145 AAAAATGCAATCTCCGACACATGGGAAGTATCGGTATTCGTGAGACTTAAGAAAA 1204  
QY 121 CCACCTCTATGACACACAGCCAAATCCACCTTCATCTGATGCGAGATTGACAAA 180  
DB 1205 CCACCTCTATGACACACAGCCAAATCCACCTTCATCTGATGCGAGATTGACAAA 1264  
QY 181 GAATGTGTGTTAGTCTTGAACAAATCTGGAAGCATGGC 220  
DB 1265 GAATGTGTGTTAGTCTTGAACAAATCTGGAAGCATGGC 1304

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (caccl)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
2 (bases 1 to 2826)  
Agnel, M., and Culouscou, J.-M.  
Direct Submission  
Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrieres, Rueil-Malmaison 92500, France

FEATURES  
source  
1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="caccl"  
5. .2749  
/gene="caccl"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD5487.1"  
/db\_xref="pid:g4585469"  
/db\_xref="gi:4585469"  
/translation="MGPKSSVFLLIHLLEGALSLSLIQLNNNGEIVVAIDPNP  
EDELIIQIKDMYQASLYLEATGKRFYFNVAIIPETWTKADYVRKLEYKNA  
DVLVAESTPPNDPEYTEMNGCKEGERIHLPDFIAGKLAEGPOGKAHEVMAH  
LRMGVFEYNDDEKFLSNGRIQAVRSAGITGVNKKCOGSGCYTRKCTFNKVTGL  
YKCEGFYLSRQTERKASIMFAOHVDSIVFECTEONNKKKAPNOKNCRSMWEV  
RDSDFKRTTPPTTPPTPSLQIGRIYCLVDKSGMATGRNLRNLOAQFL  
LQVLELSWGVAVTFDSAHVQSELIOINSQSDROTAKRLPAASGGSISCSLSRA  
FTVIRKKYPTDGESEIVLLTDEEDNTISGCFNEVKSQSAIHTVALGPSAOLELELSK  
MTGLOIYASPOVONNGILDAFGALSNGAVSORSIOLESGLTLQNSOMNGTVIY  
DSYVKDITFLITWTQPOLILMDPSOGOGFVVDKTKMAVLIQIGIAKGVWXY  
SLQASQTLTLTVTSRASNATIPITYTSKNTKSTKPSPLVYVYANIRGASPIIRA  
SVTALIESVNGKTVLELNDGAGADATKDGYSRTFTYDINGRISVYKRALGVN  
AARRVVIPOQSGALYIPGWIENDEIQNNPPPEINKDVQHRQVCFSTSSGGSFVAS  
DVNPAPIPDLFPPGOITDLKAEIHGSLINLTWAPGDYDGHGAHXYIIRISYILD  
LRDKFNESLQVNTALIPKEANSEVEFLFKNITFENGTDLFIQIADVYDKSEI  
SNIRVSLFIPQPPETPSPDETSAFCPNHINSTIPGHIILKIMMKWIGELQLSIA  
"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 100.0%; Score 220; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 3,1e-115;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTATAGTTGAATTCGTGTACAGAAACCAACCAAGAGAGCTCCAAACAGCAAAATC 60  
DB 738 CTATAGTTGAATTCGTGTACAGAAACCAACCAAGAGAGCTCCAAACAGCAAAATC 797  
QY 61 AAAAATGCAATCTCCGACACATGGGAAGTATCGGTATTCGTGAGACTTAAGAAAA 120  
DB 798 AAAAATGCAATCTCCGACACATGGGAAGTATCGGTATTCGTGAGACTTAAGAAAA 857  
QY 121 CCACCTCTATGACACACAGCCAAATCCACCTTCATCTGATGCGAGATTGACAAA 180  
DB 858 CCACCTCTATGACACACAGCCAAATCCACCTTCATCTGATGCGAGATTGACAAA 917  
QY 181 GAATGTGTGTTAGTCTTGAACAAATCTGGAAGCATGGC 220  
DB 918 GAATGTGTGTTAGTCTTGAACAAATCTGGAAGCATGGC 957

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS AF039401  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hclcal) gene,  
complete cds.

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:33:59 ; Search time 3268.17 seconds  
(without alignments)  
214.086 Million cell updates/sec

Title:	US-09-049-696-5
Perfect score:	220
Sequence:	1 CTATAGTTGAATTCGTACA.....GACAAATCTGAAGCATGGC. 220

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

```

1:  gb_ba1:*
2:  gb_ba2:*
3:  gb_om:*
4:  gb_ov:*
5:  gb_pat:*
6:  gb_ph:*
7:  gb_p11:*
8:  gb_p12:*
9:  gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*

```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	220	100.0	3311	11	AF039400	Homo sapi

2	220	100.0	2825	42	AF127036
3	218	99.1	35378	11	AF039401
4	20	9.1	15805	11	AC005666
5	20	9.1	12673	12	AB017156
6	20	9.1	142178	35	AC007114
7	19	8.6	142178	35	AC007114
8	19	8.6	1616	35	EOY17505
9	19	8.6	1765	36	EOY14933
10	18	8.2	155382	11	AC009411
11	18	8.2	40247	11	HSC00517
12	18	8.2	188768	12	AC005817
13	18	8.2	3879	12	AF078112
14	18	8.2	201175	34	AC004676
15	18	8.2	186820	34	AC006400
16	18	8.2	158262	35	AC007739
17	18	8.2	40469	35	MPAGE1
18	18	8.2	839	36	PTU71284
19	17	7.7	253505	1	AP000007
20	17	7.7	12343	2	U67505
21	17	7.7	13667	2	U67585
22	17	7.7	71522	7	AB006703
23	17	7.7	95266	7	ATP18E5
24	17	7.7	88164	7	ATP9A14
25	17	7.7	2128	7	MCU51563
26	17	7.7	37235	7	SPBC83
27	17	7.7	1217	8	AF009067
28	17	7.7	4309	9	AB018285
29	17	7.7	10000	9	AP000073
30	17	7.7	117939	9	HS21IA9
31	17	7.7	97847	9	HS874C20
32	17	7.7	31427	10	HSV521E8
33	17	7.7	106799	11	AC002538
34	17	7.7	127514	11	AC004447
35	17	7.7	175120	11	AC004687
36	17	7.7	63749	11	AC004654
37	17	7.7	108778	11	AC005194
38	17	7.7	177503	11	AC005297
39	17	7.7	187543	11	AC005678
40	17	7.7	200349	11	HSU5195
41	17	7.7	215386	11	HSU91322
42	17	7.7	88036	11	HUAC003119
43	17	7.7	250529	12	HUAE000658
44	17	7.7	24304	12	MMCD14
45	17	7.7	274433	35	AC007368

## RESULT 15

ID 071308 standard; DNA; 12284 BP.  
 AC 071308:  
 DT 25-APR-1995 (first entry)  
 DE Sequence encoding 55 kilodalton protein of hog cholera virus.  
 KW Probe; detection; identification; diagnosis; hog cholera virus;  
 OS bovine viral diarrhoea virus; HCV; BVDV; ds.  
 Synthesis.  
 FH Key  
 FT cds location/Qualifiers  
 FT 364 .12060  
 FT /tag= a  
 FT /product= 55 kDa protein of Hog Cholera Virus.  
 PN Ep-614979-A.  
 PD 14-SEP-1994.  
 PF 12-MAR-1990; 200573.  
 PR 19-MAR-1989; EP-104921.  
 PA (ALBU) AKZO NOBEL NV.  
 PI Meyers G, Ruemenapf T, Thiel H;  
 DR WPI; 94-281207/35.  
 P-PSDB; R60543.  
 PT Hog cholera virus nucleic acid and polypeptide(s) - used to  
 develop prods. for use as vaccines and for diagnosis of hog  
 cholera virus infection  
 PT Claim 1; Figure 2; 63pp; English.  
 PS Antigenic fragments of the Hog Cholera Virus (HCV) 55 kDa  
 CC polypeptide may be used to induce virus neutralising antibodies.  
 CC The antigenic fragments; cells capable of producing them and  
 CC recombinant HCV may be used in vaccines to protect animals against  
 HCV infection. A nucleic acid sequence comprising a vector nucleic  
 CC acid molecule and encoding an antigenic fragment may be used for the  
 CC diagnosis of HCV infection and can discriminate HCV from bovine  
 CC viral diarrhoea virus. The antigenic fragments can also be used in  
 CC diagnostic methods to detect the presence of HCV.  
 SQ Sequence 12284 BP: 3793 A; 2600 C; 3230 G; 2661 T;

Query Match 6.8%; Score 15; DB 1; Length 12284;

Best Local Similarity 100.0%; Pred. No. 74;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 CAGAACAAACCACA 33  
 ||||||||||||||||

DB 9991 CAGAACAAACCACA 10005

Search completed: August 5, 1999, 17:39:59  
 Job time: 6234 sec

KM 19G: Fc region binding domain; Protein A; luciferase; pUXSPARC-1;  
KW ss. Location/Qualifiers  
FH key 16..1275  
FT cds /\*tag= a  
FT cds 1302..2276  
FT cds /\*tag= b  
J02308791-A.  
PD 21-DEC-1990.  
PF 24-MAY-1989: 131195.  
PR 24-MAY-1989: JP-131195.  
PA (SEKI) SEKISUI CHEM IND KK.  
DR WPI: 91-041061/06.  
PT Polypeptide for fused protein prodn. - has binding ability to  
immunoglobulin Fc region and luciferase activity  
PS Claim 2: Fig 7: 18pp; Japanese.  
CC DNA encoding protein A and having BamHI sticky ends is ligated to  
BamHI-digested plasmid pDR720 to give pSPARC-1. A sequence  
encoding luciferase was isolated from V.harveyi DNA and ligated to  
CC BamHI-digested pUC18 to give pUX1801. A series of plasmids was  
CC derived from this plasmid, including pUX1803 which was digested  
CC with HindIII and ligated with HindIII-digested pSPARC-1. Plasmids  
CC pUXSPARC-1 and -2 were obtained. They encode fusion proteins for  
CC use as a second labelled antibody. See also Q10349 and Q10351.  
SQ Sequence 2642 BP: 761 A; 559 C; 627 G; 695 T;

Query Match 6.8%; Score 15; DB 1; Length 2642;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 CAAACCCACACACAA 38  
|||||  
DB 1189 CAAACCCACACACAA 1175

RESULT 12  
Q10351/c  
ID Q10351 standard; DNA: 2633 BP.  
AC Q10351;  
DT 15-APR-1991 (first entry)  
DE pUXSPARC-2 encoding 19G-binding/luciferase fusion protein.  
KW IGC: Fc region binding domain; Protein A; luciferase; pUXSPARC-2;  
KW ss. Location/Qualifiers  
FH key 16..1266  
FT cds /\*tag= a  
FT cds 1293..2267  
FT cds /\*tag= b  
J02308791-A.  
PD 21-DEC-1990.  
PF 24-MAY-1989: 131195.  
PR 24-MAY-1989: JP-131195.  
PA (SEKI) SEKISUI CHEM IND KK.  
DR WPI: 91-041061/06.  
PT Polypeptide for fused protein prodn. - has binding ability to  
immunoglobulin Fc region and luciferase activity  
PS Claim 3: Fig 8: 18pp; Japanese.  
CC DNA encoding protein A and having BamHI sticky ends is ligated to  
CC BamHI-digested plasmid pDR720 to give pSPARC-1. A sequence  
CC encoding luciferase was isolated from V.harveyi DNA and ligated to  
CC BamHI-digested pUC18 to give pUX1801. A series of plasmids was  
CC derived from this plasmid, including pUX1803 which was digested  
CC with HindIII and ligated with HindIII-digested pSPARC-1. Plasmids  
CC pUXSPARC-1 and -2 were obtained. They encode fusion proteins for  
CC use as a second labelled antibody. See also Q10349 and Q10350.  
SQ Sequence 2633 BP: 755 A; 561 C; 627 G; 690 T;

Query Match 6.8%; Score 15; DB 1; Length 2633;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 CAAACCCACACACAA 38  
|||||  
DB 1180 CAAACCCACACACAA 1166

RESULT 13  
Q51348/c  
ID Q51348 standard; DNA: 840 BP.  
AC Q51348;  
DT 16-MAY-1994 (first entry)  
DE Aspergillus promoter sequence.  
KW Protein production; high levels; ss.  
OS Aspergillus nidulans FGSC No. A89.  
PN J05268964-A.  
PD 19-OCT-1993.  
PF 30-MAR-1992: 074092.  
PR 30-MAR-1992: JP-074092.  
PA (MEIP) MEIJIRI MILK PROD CO LTD.  
DR WPI: 93-364289/46.  
PT Promoter of aspergillus used to produce animal protein - has  
PS sequence of 480 base pairs  
PS Example: Page 6: 8pp; Japanese.  
CC The sequence is that of an Aspergillus promoter which can be used  
CC to produce high levels of animal protein.  
SQ Sequence 840 BP: 221 A; 215 C; 181 G; 223 T;

Query Match 6.8%; Score 15; DB 1; Length 840;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ACAACAAAGAGCTC 45  
|||||  
DB 494 ACAACAAAGAGCTC 480

RESULT 14  
Q50632  
ID Q50632 standard; cDNA: 1860 BP.  
AC Q50632;  
DT 02-JUN-1994 (first entry)  
DE Brain factor-2.  
KW Brain factor; BF-1; BF-2; BF-3; DNA binding domain;  
KW brain transcription factor; diagnosis; tumour; cancer; probe;  
KW telencephalon; ss.  
OS Rattus rattus.  
PN W09333430-A.  
PD 25-NOV-1993.  
PF 30-APR-1993: U04102.  
PR 13-MAY-1992: US-882292.  
PA (SLOK) SLOAN KETTERING INST CANCER.  
DR IAL E, TAO W;  
DR WPI: 93-386481/48.  
PT Isolated, animal nucleic acids encoding brain transcription  
PT factors - useful for diagnosis and treatment of abnormal brain  
PT factor synthesis in tumour tissue from animals and utilised as  
PT probe  
PS Claim 2: Page 62-63: 96pp; English.  
CC The brain factors are transcription factors. Abnormal expression of  
CC BF-1 in telencephalon-derived tissue or tumour tissue can be  
CC diagnosed. BF DNA or protein can be used to correct defective  
CC synthesis of BF.  
SQ Sequence 1860 BP: 402 A; 543 C; 545 G; 370 T;

Query Match 6.8%; Score 15; DB 1; Length 1860;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 AAACAGCAAAATCA 61  
|||||  
DB 1453 AAACAGCAAAATCA 1467

QY 47 AACACAGCAAAATCAA 62  
 |||||||  
 DB 271 AACACAGCAAAATCAA 286

## RESULT 8

V75326  
 ID V75326 standard; DNA: 462 BP.  
 AC V75326;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #1015.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PE 07-JAN-1997; 100117.  
 PF 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 PI Rosen CA;  
 DR WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1; Page 1794; 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 SO Sequence 462 BP; 206 A; 75 C; 77 G; 104 T;

Query Match 7.3%; Score 16; DB 1; Length 462;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 48 AACACAGCAAAATCAA 63  
 |||||||  
 DB 174 AACACAGCAAAATCAA 189

RESULT 9  
 X30886/c  
 ID X30886 standard; DNA: 528 BP.  
 AC X30886;  
 DT 20-MAY-1999 (first entry)  
 DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:163.  
 KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;  
 KW streptococcal infection; pneumococcal; ss.  
 OS Streptococcus pneumoniae.  
 PN WO9737026-A1.  
 PD 09-OCT-1997.  
 PF 01-APR-1997; 005306.  
 PR 22-AUG-1996; US-025788.  
 PR 02-APR-1996; US-014690.

PA (SMK ) SMITHKLINE BEECHAM CORP.  
 PA (SMK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,  
 PI Stodola RK;  
 DR WPI: 97-503111/46.  
 DR P-PSDB: Y11304.  
 PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
 PT vaccines, drug screening, etc  
 PS Claim 5; Page 153; 354pp; English.  
 CC X30724 to X30946 represent genomic DNA sequences isolated from  
 CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
 CC encode the novel proteins given in Y1114 to Y1167. The proteins,  
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
 CC streptococcal infections and in assays for identifying compounds that  
 CC inhibit or activate the activity of the proteins. The antagonists can  
 CC be used to treat an individual having need to inhibit a bacterial  
 CC protein. Vectors expressing the proteins can be used to induce a  
 CC protective immune response in mammals.  
 SO Sequence 528 BP; 158 A; 88 C; 109 G; 173 T;

Query Match 7.3%; Score 16; DB 1; Length 528;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 AACACAGCAAAATCAA 62  
 |||||||  
 DB 255 AACACAGCAAAATCAA 240

RESULT 10  
 Q06001  
 ID Q06001 standard; DNA: 12284 BP.  
 AC Q06001;  
 DT 16-JAN-1991 (first entry)  
 DE Sequence encoding protein characteristic of hog cholera virus (HCV).  
 KW Vaccine; parvovirus; swine influenza virus.  
 OS Hog cholera virus.  
 FH Key Location/Qualifiers  
 FT cds 364..12057  
 FT /\*tag= a  
 PN EP-389034-A.  
 PD 26-SEP-1990.  
 PF 12-MAR-1990; 200573.  
 PR 19-MAR-1989; EP-104921.  
 PR 12-MAR-1990; EP-200573.  
 PA (ALKU ) AKZO NV.  
 PI Meyers G, Rumenapf T, Thiel HJ;  
 DR WPI: 90-291979/39.  
 DR P-PSDB: R06996.  
 PT New hog cholera virus vaccine and diagnostic - comprises nucleic  
 PT acid sequence of poly-peptide characteristic of hog cholera virus  
 PS Claim 2; Fig 2; 29pp; English.  
 CC Gene product may be used to provide a vaccine and Abs for diagnosis  
 CC of hog cholera viral infection in pigs.  
 SO Sequence 12284 BP; 3793 A; 2601 C; 3228 G; 2662 T;

Query Match 6.8%; Score 15; DB 1; Length 12284;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 19 CAGACACAAACACACA 33  
 |||||||  
 DB 9991 CAGACACAAACACACA 10005  
 RESULT 11  
 Q10350/c  
 ID Q10350 standard; DNA: 2642 BP.  
 AC Q10350;  
 DT 15-APR-1991 (first entry)  
 DE pLUXSPAc-1 encoding IgG-binding/luciferase fusion protein.

PN W09818931-A2.  
 PD 07-MAY-1998.  
 PF 30-OCT-1997; 019588.  
 PR 31-OCT-1996; US-029960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barrach SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
 PI Kunsch CA, Rosen CA;  
 PI WPI: 98-272225/24.  
 PT Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 PS Claim 1: Page 1078-1080; 1409pp; English.  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridize to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid molecules whose nucleotide  
 CC sequence is homologous to amplification primers derived from the  
 CC fragment of the S. pneumoniae genome to prime the amplification and  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating  
 CC fragments of the S. pneumoniae genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for S. pneumoniae.  
 CC Sequence 4104 BP; 1264 A; 711 C; 924 G; 1205 T;  
 SQ

Query Match 7.3%; Score 16; DB 1; Length 4104;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 AACACAGCAAAATCAA 62  
 DB 3185 AACACAGCAAAATCAA 3170

RESULT 6  
 V43006/c  
 ID V43006 standard; DNA; 2130 BP.  
 AC V43006:  
 DT 09-NOV-1998 (first entry)  
 DE Streptococcus pneumoniae polypeptide coding region.  
 KW Polypeptide; ORF; open reading frame; infection; bacterial;  
 KW streptococcal; bacteremia; diagnosis; prophylaxis; ds.  
 OS Streptococcus pneumoniae.  
 FH Key Location/Qualifiers  
 FT CDS 1915..2097  
 FT /\*tag= a  
 FT /note= "polypeptide"  
 PN W09823631-A1.  
 PD 04-JUN-1998.  
 PF 24-NOV-1997; 021976.  
 PR 27-NOV-1996; US-031879.  
 PA (SMIR ) SMITHKLINE BEECHAM CORP.  
 PI (SMIR ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,  
 PI Reid RH, Zaros PN;  
 PI WPI: 98-322654/28.  
 DR P-PSDB: W62731.  
 PT Streptococcus pneumoniae polynucleotides - useful for developing  
 PT products for diagnosis, prevention and treatment of infections e.g.  
 PT pneumonia, bacteremia, meningitis or endocarditis

PS Claim 1: Page 122-123; 181pp; English.  
 CC The sequence is that of a Streptococcal polypeptide coding region.  
 CC The polypeptide can potentially be used for the diagnosis and  
 CC prevention of bacterial infections, especially SP infection.  
 CC It may be used for the treatment of diseases such as otitis media,  
 CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural  
 CC empyema, endocarditis or infection of the cerebrospinal fluid.  
 CC Sequence 2130 BP; 639 A; 385 C; 498 G; 608 T;  
 SQ

Query Match 7.3%; Score 16; DB 1; Length 2130;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 AACACAGCAAAATCAA 62  
 DB 1369 AACACAGCAAAATCAA 1354

RESULT 7  
 T98762 standard; DNA; 529 BP.  
 ID T98762:  
 AC T98762:  
 DT 10-NOV-1998 (first entry)  
 DE DNA encoding a S. pneumoniae PFS protein.  
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
 KW immunological response; inoculation; antibody production; inhibitor;  
 KW T cell immune response; antimicrobial compound; bacterial adhesion;  
 KW extracellular matrix protein; protein-mediated cell invasion; wound;  
 KW pathogenesis; ss.  
 OS Streptococcus pneumoniae.  
 FH Key Location/Qualifiers  
 FT CDS complement (40..234)  
 FT /\*tag= a

PN W09743303-A1.  
 PD 20-NOV-1997.  
 PF 14-MAY-1997; 007950.  
 PR 14-MAY-1996; US-0117670.  
 PA (SMIR ) SMITHKLINE BEECHAM CORP.  
 PI (SMIR ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,  
 PI Stodola RK;  
 PI WPI: 98-008793/01.  
 DR P-PSDB: W38720.  
 PT Novel Streptococcus pneumoniae proteins and related DNA -- useful for  
 PT diagnosing anti-microbial agents for treatment of bacterial  
 PT infections  
 PS Claim 4: Page 244; 483pp; English.  
 CC This sequence encodes a Streptococcus pneumoniae protein that (based on  
 CC homology with an Escherichia coli protein) is a PFS protein, and  
 CC represents a DNA sequence of the invention. The DNA sequences were  
 CC isolated from Streptococcus pneumoniae strain 0100993 (NCIMB 40794). The  
 CC Streptococcus pneumoniae proteins of the invention can be used to  
 CC identify compounds which interact with and inhibit or activate the  
 CC activity of the proteins. Antagonists can be used to treat diseases  
 CC caused by S. pneumoniae proteins, through genetic immunisation. They can  
 CC also be used to induce an immunological response in a mammal by  
 CC inoculation with the S. pneumoniae proteins or delivery of the encoding  
 CC nucleic acids in a vector adequate to produce antibody and/or T cell  
 CC immune responses to protect the animal from disease. The proteins can  
 CC also be used to identify antimicrobial compounds which are capable of  
 CC inhibiting their bioactivity. In particular the proteins of the invention  
 CC can be used to prevent adhesion of bacteria to mammalian extracellular  
 CC matrix proteins on in-dwelling devices or in wounds, to block progression of  
 CC mediated mammalian cell invasion, and to block the normal progression of  
 CC pathogenesis in infections initiated other than by the implantation of  
 CC in-dwelling devices or other surgical techniques.  
 CC Sequence 529 BP; 173 A; 111 C; 89 G; 156 T;  
 SQ

Query Match 7.3%; Score 16; DB 1; Length 529;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



OY 169 AGATTGACAAAGAAAT 185  
 DB 63555 AGATTGACAAAGAAAT 63539

# RESULT 2

V21209\_09/c  
 Continuation (10 of 17) of V21209 from base 900001 (Methanococcus jannaschii circular ch  
 WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match  
 Best Local Similarity 7.7%; Score 17; DB 1; Length 110000;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 AAAATCAAAATGCAAT 71  
 DB 70170 AAAATCAAAATGCAAT 70154

# RESULT 3

X13711/c  
 ID X13711 standard; DNA: 471 BP.

DE 19-MAR-1999 (first entry)  
 DE Enterococcus faecalis genome contig SEQ ID NO:774.  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
 KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN WO980555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; US-066009.

PR 14-NOV-1997; US-046009.

PR 06-MAY-1997; US-044031.

PR 16-MAY-1997; US-046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI: 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.

SO Sequence 471 BP; 139 A; 80 C; 80 G; 170 T;

Query Match  
 Best Local Similarity 7.7%; Score 17; DB 1; Length 471;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 141 CCACCAATCCACCTT 157  
 DB 375 CCACCAATCCACCTT 359

# RESULT 4

V86909/c  
 ID V86909 standard; cDNA: 376 BP.

AC V86909; (first entry)  
 DE 27-APR-1999 (first entry)  
 DE EST clone B48.

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

PN WO9845435-A2.

PD 15-OCT-1998.

PR 10-APR-1998; US-835913.

PR 10-APR-1997; US-835913.

PA (GENE) GENETICS INST INC.

PI Agostino MJ, Jacobs ER, McCoy JM, Merberg D,

PI Racie LA, Spaulding V, Treacy M;

DR WPI: 99-070076/06.

PT New polynucleotides encoding human secreted proteins - derived from

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,

PT ovary, pituitary, retina and colon cDNA libraries

PS Claim 1; Page 388; 633pp; English.

CC This sequence represents an expressed sequence tag (EST), and is a

CC polynucleotide of the invention. The polynucleotides of the invention are

CC all secreted EST sequences isolated from a variety of human tissue

CC sources. The EST sequences and proteins encoded by them are predicted to

CC have useful biological activities which would make them suitable for

CC treating, preventing or ameliorating medical conditions in humans and

CC animals, although no supporting data is given. Suggested activities

CC include nutritional activity, immune stimulating or suppressing activity,

CC haematopoiesis regulating activity, tissue growth activity,

CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition

CC activity. The EST sequences are also stated to be useful for gene

Sequence 376 BP; 101 A; 58 C; 71 G; 146 T;

Query Match  
 Best Local Similarity 7.7%; Score 17; DB 1; Length 376;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 AACAAACCAACACAA 38  
 DB 198 AACAAACCAACACAA 182

# RESULT 5

V52302/c  
 ID V52302 standard; DNA: 4104 BP.

AC V52302; (first entry)  
 DE 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:169.

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

Mon Aug 9 13:27:07 1999

us-09-049-696-5\_1.rng

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:39:52 ; Search time 650.92 Seconds  
(without alignments)  
84.561 Million cell updates/sec

Title: US-09-049-696-5

Perfect score: 220

Sequence: 1 CTAATGTCATCTGTGACA.....GACAAATCTGAGACATGCC 220

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	7.7	110000	1 V21209_00	Methanococcus jann
2	17	7.7	110000	1 V21209_09	Continuation (10 o
3	17	7.7	471	1 X13711	Enterococcus faeca
4	17	7.7	376	1 V86909	EST clone BA88. Ne
5	16	7.3	4104	1 V52302	Streptococcus pneu
6	16	7.3	2130	1 V43006	Streptococcus pneu
7	16	7.3	529	1 T98762	DNA encoding a S.
8	16	7.3	462	1 V75326	Streptococcus pneu
9	16	7.3	528	1 X30886	Sequence encoding
10	15	6.8	12284	1 O06001	pluxSPAC-1 encodi
11	15	6.8	2642	1 O10350	pluxSPAC-2 encodi
12	15	6.8	2633	1 Q10351	Aspergillus promot
13	15	6.8	840	1 O51348	Brain factor-2. Is
14	15	6.8	1860	1 O50632	Sequence encoding
15	15	6.8	12284	1 O71308	Rat SIII 110 kDa s
16	15	6.8	3501	1 O90529	Human GRB-10 cDNA
17	15	6.8	2556	1 T03197	Human GRB-10 cDNA
18	15	6.8	2403	1 T11081	V.harveyi luxAB ge
19	15	6.8	4472	1 T05406	Human cell cycle g
20	15	6.8	14086	1 T28773	Human calcium sens
21	15	6.8	14042	1 T28774	Human placental ca
22	15	6.8	14080	1 T28775	Human kidney calci
23	15	6.8	14044	1 T28776	Human parathyroid
24	15	6.8	2520	1 T09941	Another earl gene p
25	15	6.8	12284	1 T29591	Hog cholera virus.
26	15	6.8	3501	1 T59995	Rat RNA polymerase
27	15	6.8	6803	1 T85473	Genomic hMET seque
28	15	6.8	14086	1 V05994	Human calcium prot
29	15	6.8	14042	1 V05995	Human placenta cal
30	15	6.8	14080	1 V05996	Human kidney calci
31	15	6.8	14044	1 V05997	Human parathyroid
32	15	6.8	1881	1 V09214	Crkl cDNA sequence
33	15	6.8	6803	1 V16305	Genomic DNA encodi
34	15	6.8	2167	1 V20678	Sf9 alpha-mannosid
35	15	6.8	2167	1 V20678	Sf9 alpha-mannosid
36	15	6.8	1404	1 V27212	Stimulador of iron
37	15	6.8	2007	1 V40754	C. felis esterase,
38	15	6.8	2007	1 V40755	C. felis esterase,
39	15	6.8	1540	1 V40756	C. felis esterase,
40	15	6.8	1584	1 V40757	C. felis esterase,
41	15	6.8	15229	1 V18276	RSV isolate 18537
42	15	6.8	486	1 V1295	E. coli J96 pathog
43	15	6.8	7001	1 V52221	Streptococcus pneu

c 44 15 6.8 110000 1 V21209\_02  
45 15 6.8 1514 1 X22018

## ALIGNMENTS

RESULT	1	LOCUS	V21209	Accession	V21209
V21209_00/c					
WP Sequence split into 17 fragments					
WP Fragment Name	Begin	End			
WP V21209_00	1	110000			
WP V21209_01	100001	210000			
WP V21209_02	200001	310000			
WP V21209_03	300001	410000			
WP V21209_04	400001	510000			
WP V21209_05	500001	610000			
WP V21209_06	600001	710000			
WP V21209_07	700001	810000			
WP V21209_08	800001	910000			
WP V21209_09	900001	1010000			
WP V21209_10	1000001	1110000			
WP V21209_11	1100001	1210000			
WP V21209_12	1200001	1310000			
WP V21209_13	1300001	1410000			
WP V21209_14	1400001	1510000			
WP V21209_15	1500001	1610000			
WP V21209_16	1600001	1664976			
WP V21209_17	1664976				
ID	V21209	standard: DNA; 1664976 BP.			
AC	V21209;				
DT	10-NOV-1998 (first entry)				
DE	Methanococcus jannaschii circular chromosome.				
KW	Methanococcus jannaschii; methanogenic archaeon; circular chromosome;				
OS	genome; autotrophic; extrachromosomal element; identification; ds.				
PN	Methanococcus jannaschii.				
PD	26-FEB-1998.				
PF	22-AUG-1997; U14900.				
PR	22-AUG-1996; US-024428.				
PA	(GENO-) INST GENOMIC RES.				
PA	(UNIT-) UNIV JOHNS HOPKINS SCHOOL MEDICINE.				
PA	(BULT) BULT HO, Venter JC, White OR, Moese CR;				
DR	WPI; 98-169145/15.				
PT	Complete genome sequence of methano-genic archaeon. Methanococcus				
PT	jannaschii - useful in identification of M. jannaschii genome				
PT	fragment				
PS	Claim 13; Page 152-585; 614pp; English.				
CC	The present sequence represents the complete 1.66-megabase pair genome				
CC	sequence of the Methanococcus jannaschii open reading frames from the				
CC	present invention describes M. jannaschii open reading frames from the				
CC	genome sequence. The invention also describes a computer based system				
CC	for identifying fragments of the M. jannaschii genome that are				
CC	homologous to target nucleotide sequences, comprising: (a) data storage				
CC	means comprising the nucleotide sequence of the 1664976, 58407 or 16550				
CC	bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at				
CC	least 99.9% identical to it; (b) search means for comparing a target				
CC	sequence to the nucleotide sequence of the data storage means to				
CC	identify a homologous sequence, and (c) retrieval means for obtaining				
CC	the homologous sequence. The method, which is based on whole genome				
CC	random sequencing of an autotrophic archaeon M. jannaschii, the genome				
CC	of which consists of 3 physically distinct elements, a large circular				
CC	chromosome (the 1664976 bp sequence given in V21209), a large circular				
CC	extra-chromosomal element (the 58407 bp sequence given in V21210), and a				
CC	small circular extra-chromosomal element (the 16550 bp sequence given in				
CC	V21211), can be used in the identification of M. jannaschii genome				
CC	fragment.				
SQ	Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T;				

Query Match 7.7%; Score 17; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 6.8; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0

Continuation (3 of  
Partial mouse MED1

---

This Page Blank (uspto)

	Query Match	13.0%;	Score 28.6;	DB 1;	Length 582;	
	Best Local Similarity	61.3%;	Pred. NO. 10;			
	Matches 46;	Conservative	0; Mismatches	29; Indels	0; Gaps	0;
QY	20 AGAAGAAACCCATCACTCAGAGTCCAAACAGCAAAATGCAATTCCTCCGAG	79				
Dd	459 AGAAGACTCCACCACCAAAGAAGATTCTATGATGTAAAAACCAAGACAAGCGTTTCTTA	518				
OY	80 CACATGGGAAGTGAT	94				
Dd	519 CATACCGAAGAATAAT	533				

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664576

Query Match	13.0%	Score 28.6	DB 1	Length 64576
Best Local Similarity	54.2%	Pred No 36		
Matches	58	Mismatches	49	Indels 0
				Gaps 0

Oy 78 AGCACATGGAGTGCATCCGATTCGTAGACCTTAAGAAAAACCAC 124  
 || ||| || | ||| ||| | | ||| ||  
 Db 41725 TTTAACCTAAAGAAATTAGAGAAATATGAGGAGATAATTGAAATAAC 41679

```

RESULT 15
T83805/C
ID T83805 standard; DNA: 720 BP.
AC T83805;
DT 16-JUL-1998 (first entry)
DE DNA encoding a Staphylococcus aureus protein of unknown function.
KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW Staphylococcal gene; regulatory element; bacterial gene expression;
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome
KW toxic shock syndrome; ss.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT CDS complement(25..303)
FT /-tag- a
PN MO9730070-A1.
PD 21-AUG-1997.
PF 19-FEB-1997; U02318.
PR 20-FEB-1996; U5-011888.
PI (SMIT) SMITHKLINE BEECHAM CORP.
PA Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR P-PSDB: W27842.

```

FT Novel polypeptide(s) from *Staphylococcus aureus* strain WCH29 - used  
PT to isolate antimicrobial compounds, and in vaccines against *S.*  
PT *aureus* infection  
PS Claim 9; Pages 683-684; 989pp; English.  
CC The present sequence encodes a *Staphylococcus aureus* protein of  
CC unknown function. The present sequence was isolated from a  
CC library of clones of *S. aureus* WCH 29 in *Escherichia coli*. The DNA  
CC sequence can be used in the construction of ribozymes and antisense  
CC sequences to control the expression of *Staphylococcus* genes. The DNA  
CC sequence is also useful as a source of regulatory elements for the  
CC control of bacterial gene expression. The encoded protein may be used  
CC to produce vaccines to enable a host to produce specific antibodies  
CC with antibacterial action. These vaccines and antibodies would protect  
CC a host against invasion by *S. aureus*, and conditions relating to  
CC *Staphylococcus* infection, e.g. *Staphylococcal* food poisoning, scalded  
CC skin syndrome, and toxic shock syndrome.  
SQ Sequence 720 BP; 177 A; 116 C; 122 G; 294 T;

[illegible]

Search completed: August 6, 1999, 04:02:32  
Job time: 1504 sec

PE 18-SEP-1998; U19746.  
PR 19-SEP-1997; US-059401.  
PA (ORTH ) ORTHO-MCNEIL PHARM INC.  
PI Chamberlain H, Farrell F, Galindo J, Huvar A, Johnson D,  
PJ Joffe L, Patel L;  
DR WPI: 99-229542/19.  
P-PSDB: W98105.  
PT New ras carboxyl terminal processing protein useful for identifying  
PT modulators useful in therapy.  
PS Claim 2; Fig 1; 52pp; English.  
CC This cDNA sequence encodes a novel human protein (see W98105)  
CC that functions as a ras carboxyl terminal processing protein, and  
CC optionally as a ras processing enzyme (i.e. ras protease). The  
CC cDNA was cloned from a human colorectal adenocarcinoma cDNA  
CC library following a homology search using yeast RCEL, the protease  
CC responsible for ras processing in yeast. The cDNA has been  
CC expressed in recombinant host cells which produce active  
CC recombinant protein. The recombinant protein, and recombinant  
CC host cells are utilised in a method for identifying modulators  
CC of the enzyme activity, useful for treating a condition mediated  
CC by activated ras protein. Inhibition of the human ras protease  
CC should be efficacious for cancer treatment. Recombinant DNA  
CC molecules, and portions of them, are useful for isolating  
CC homologous sequences, identifying and isolating genomic  
CC equivalents, and identifying, detecting or isolating mutant  
CC forms of the ras carboxyl terminal processing protein DNA.  
S0 Sequence 1549 BP; 299 A; 440 C; 433 G; 374 T;  
  
Query Match 13.1%; Score 28.8; DB 1; Length 1549;  
Best Local Similarity 54.8%; Pred. No. 11;  
Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;  
  
OY 3 ATAGTGAATTCGTACAGACAAACCAACAAAGAGCTCCAAACAGCAAAATCAA 62  
DB 1437 AGATTAAATGCTGTAACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1496  
OY 63 AATGCAATCTCCGAACGACATGATCGTGGATTCGA 106  
DB 1497 AAAAAAAGAAAAAAGAAAAAAGGCGCGCTCGCATCTAGA 1540  
  
RESULT 12  
T67702  
ID T67702 standard; DNA: 462 BP.  
AC T67702;  
DT 15-JUL-1997 (first entry)  
DE H. pylori transmembrane protein ORF 5993958.aa.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
KW binding compound; bacterium; life cycle; activator; bacteriophage; inhibitor;  
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.  
OS Helicobacter pylori.  
FH Key  
FT cds  
FT 1. 462  
FT /tag- a  
FT /note- "no stop codon given"  
P-PSDB: W20552.  
PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
PS Claim 27; Page 311, 1481pp; English.  
CC This sequence encodes a H. pylori protein likely to contain two  
CC membrane spanning regions.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
S0 Sequence 582 BP; 203 A; 92 C; 126 G; 161 T;

CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
S0 Sequence 462 BP; 164 A; 75 C; 91 G; 132 T;  
  
Query Match 13.0%; Score 28.6; DB 1; Length 462;  
Best Local Similarity 61.3%; Pred. No. 9.4;  
Matches 46; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
  
OY 20 AGAACAAACCAACAAAGAGCTCCAAACCAACCAAAATGCAATCTCCGAG 79  
DB 339 AGAAGAGTCCACCAACAAAGAGATCTTAATGATGAACCAACCAAGACAGTTCTAA 398  
OY 80 CACATGGAGAGTGAAT 94  
DB 399 CATTCACCAAGAAAT 413  
  
RESULT 13  
T68030  
ID T68030 standard; DNA: 582 BP.  
AC T68030;  
DT 15-JUL-1997 (first entry)  
DE H. pylori transmembrane protein ORF 07ge20415orf19.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacterium; life cycle; activator;  
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
KW diagnosis; ds.  
OS Helicobacter pylori.  
FH Key  
FT cds  
FT 1. 582  
FT /tag- a  
FT /note- "no stop codon given"  
P-PSDB: W20777.  
PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
PS Claim 27; Page 856; 1481pp; English.  
CC This sequence encodes a H. pylori protein likely to contain two  
CC membrane spanning regions.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
S0 Sequence 582 BP; 203 A; 92 C; 126 G; 161 T;



[illegible]

OS Brassica napus var. Rafel.  
FH Key Location/Qualifiers  
FT cds /tag= a  
FT 44..1199  
FT 44..1199  
FT signal\_peptide /label= stearyl-ACP-desaturase  
FT 44..263  
FT /tag= b  
FT 264..1199  
FT mat\_peptide /tag= c  
FT  
FT NL002130-A.  
PN 16-APR-1992.  
PD 28-SEP-1990: 002130.  
PE 28-SEP-1990: NL-002130.  
PR (TFME-) STICHTING TECH WETENSCHAPPEN.  
PA WPI: 92-157514/19.  
DR P-PSDB: R23793.  
DR Cruciferous stearyl-ACP-desaturase coding sequences - for  
PT modifying lipid biosynthesis in plants, esp. oilseed rape  
PS The sequence, obt. from clone PDEST, does not contain the poly-A-  
CC tail present in other clones (reason unknown).  
CC Disclosure: Fig 1(a-b): 31pp: Dutch  
CC stearyl-ACP-desaturase catalyses the conversion of stearyl-ACP  
CC into oleoyl-ACP. Vectors contg. the sequence have been used to  
CC transform petunia and oilseed rape plants using Agrobacterium  
CC tumefaciens, resulting in altered lipid biosynthesis and  
CC modification of lipid composition of the plant.  
SQ Sequence 1358 BP: 357 A; 293 C; 311 G; 397 T;

Query Match 13.5%; Score 29.8; DB 1; Length 1358;  
Best Local Similarity 56.7%; Pred. No. 5.6; Indels 0; Gaps 0;  
Matches 55; Conservative 0; Mismatches 42;

OY 18 ACAGAACAAACCAACCAAGAGCTCCAAACCAACCAAAATCAAAATGCAATCCGA 77  
DB 1295 ACAAACTGAAACATACATATATCTCTTAAAGGAAAAAATAACAAAGCCGA 1236  
OY 78 AGCAGATGGAGAGTGCCTGATCTGAGACTTAA 114  
DB 1235 GGAGGACATATGCTATCTCGATCAGAGACCCCTTA 1199

RESULT 6  
T08711/C  
ID T08711 standard: cDNA: 3447 BP.  
AC T08711:  
DT 25-MAY-1996 (first entry)  
DE Sak-a serine-threonine kinase cDNA.  
KM Sak-a; serine-threonine kinase; agonist; antagonist;  
KM proliferative disease; cancer; tumour; antisense; transgenic animal;  
KW therapy; ss.  
OS Mus musculus.  
FH Key Location/Qualifiers  
FT cds 206..23983  
FT /note= a  
FT /note= "T bases may also be U"  
FT misc\_difference 2767  
FT /tag= b  
FT /note= "base n at position 2767 is unclear in the  
FT specification: encoded amino acid is Ala"

CA2150789-A.  
PN 03-DEC-1995.  
PD 01-JUN-1995: 150789.  
PE 02-JUN-1994: US-252955.  
PR (MOUN) MOUNT SINAI HOSPITAL CORP.  
PA Dennis JW, Fode C, Heffernan M;  
DR WPI: 96-129817/14.  
DR P-PSDB: R92177.  
DR Nucleic acid encoding Sak serine-threonine kinase - useful for  
PT identifying modulators potentially useful in treatment or prevention  
PT of proliferative disease.  
PS Claim 5: Page 50-55: 73pp: English.  
CC A cDNA sequence (T08711) encodes sak-a (R92177), a novel

CC serine/threonine kinase associated with mitotic and meiotic cell  
CC division. It was isolated from a murine lymphoid cell (J333n25)  
CC cDNA library established in CHO cells by selection of wheatgerm  
CC agglutinin-resistant clones. Another cDNA clone (T08712) encoded  
CC the sak-b isoform (R92214). Sak-a and sak-b are probably  
CC alternatively spliced forms of the gene. The sak gene can be used  
CC for prodn. of recombinant SAK, as a probe in the diagnosis of  
CC proliferative disorders or, in antisense form, may be used to treat  
CC such disorders.  
SQ Sequence 3447 BP: 1061 A; 690 C; 754 G; 941 T;

Query Match 13.5%; Score 29.8; DB 1; Length 3447;  
Best Local Similarity 52.9%; Pred. No. 7.2; Indels 0; Gaps 0;  
Matches 64; Conservative 0; Mismatches 57;

OY 16 GTACAGACAAACCAACCAAGAGCTCCAAACCAACCAAAATCAAAATGCAATCTCC 75  
DB 3271 GTAAAGAAATTAACAAACCAACCAAGCATACACTTGTTCATACAAACATGGA 3212  
OY 76 GAAGCAGATGGAGAGTGCCTGATCTGAGACTTAAAGAAACCACTCTAGACAA 135  
DB 3211 GCACAACTTGAAGATTAACATGCTTGAGCAAAATTGATGCTTACCTTATGACCA 3152  
OY 136 C 136  
DB 3151 C 3151

RESULT 7  
T68954/C  
ID T68954 standard: cDNA: 519 BP.  
AC T68954:  
DT 05-AUG-1997 (first entry)  
DE Blackcurrant PRIB3 cDNA clone.  
KM Blackcurrant; fruit-specific promoter; PRIB3 gene;  
KW transgenic plant; ss.  
OS Ribes nigrum strain Ben Alder.  
FH Key Location/Qualifiers  
FT cds 45..242  
FT /tag= a  
FT WO9717452-A1.  
PN 15-MAY-1997.  
PD 04-NOV-1996: E04807.  
PE 03-NOV-1995: GB-022558.  
PR (SMIK) SMITHKLINE BEECHAM PLC.  
PA Brennan RM, Taylor MA, Woodhead MR;  
PI WPI: 97-281041/25.  
DR P-PSDB: W17051.  
DR New isolated promoters from blackcurrant fruit - used for driving  
PT fruit-specific expression of DNA sequences in transgenic  
PT blackcurrant and other non-climacteric fruit  
PS Claim 5: Page 25: 66pp: English.  
PS Isolated cDNA sequences (T68953-57) correspond to genes which  
CC exhibit differential expression in blackcurrant fruit during the  
CC ripening period of fruit development. They were isolated from  
CC a cDNA library constructed from the green/red stage of ripening  
CC by differential screening with probes from this stage and from  
CC green fruit. PRIB3 encodes a polypeptide (W17051) which shows  
CC a high degree of similarity to group one metallothioneins. RIB3  
CC expression levels were detected in roots, leaves and stems of  
CC plants that had borne fruit (there was no detectable expression  
CC in plants that had not borne fruit).  
SQ Sequence 519 BP: 154 A; 85 C; 125 G; 155 T;

Query Match 13.5%; Score 29.8; DB 1; Length 519;  
Best Local Similarity 50.3%; Pred. No. 4.3; Indels 0; Gaps 0;  
Matches 73; Conservative 0; Mismatches 72;

OY 2 TAAAGTGAATCTGTACAGAAACCAACCAAGAGCTCCAAACCAAGCAAAATCA 61



```
FT      /*tag= a
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
PN      EP-786519-A2.
PD      30-JUL-1997.
PF      07-JAN-1997; 100117.
PR      05-JAN-1996; US-009861.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunesh CA,
PI      Rosen CA:
PI      WPI: 97-374922/35.
PT      Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT      stored on computer readable medium and used in the production of
PT      anti-S.aureus vaccines
PS      Claim 1: Page 1374-1375: 3271pp; English.
CC      This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC      of the invention. The DNA sequences are recorded on a computer readable
CC      medium, preferably selected from a floppy or hard disk, random access
CC      memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC      the S.aureus DNA sequences allows putative functions to be assigned so
CC      that protein-encoding or regulatory regions of commercial, therapeutic or
CC      industrial importance can be obtained. Specifically, sequences which are
CC      likely to encode antigens have been identified and these polypeptides can
CC      be used in a vaccine composition against S.aureus infection. The
CC      polypeptides can also be used in a kit for the immunodetection of
CC      S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC      including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC      skin and surgical wound infections, scalded skin syndrome, toxic shock
CC      syndrome, etc. Organisms transformed with the DNA sequences can be used
CC      for recombinant production of the polypeptides. The new DNA sequences
CC      (and their fragments) are useful as primers or probes for isolating
CC      homologues of any of the S.aureus DNA sequences contained on the
CC      computer readable medium.
SQ      Sequence 1814 BP; 510 A; 350 C; 239 G; 651 T;

Query Match 14.6%; Score 32.2; DB 1; Length 1814;
Best Local Similarity 56.0%; Pred. No. 1.2;
Matches 61; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 9 GAATCTGTACAGAACCAACCAAGAGCTCCAAACCAACCAATCAAAATGCG 68
Db 1068 GAATCCGATTTATTACAAAGCTTAATGAAAGTAATTAATGAAACATCATGAT 1029

Qy 69 AATCTCCAGACACATGGAGATCCGTGATTCGTGAGACTTTAGA 117
Db 1028 TACCAACGACCTAGATGAAGTGTAGTATTTACGAAGGTAAAA 980

RESULT 3
X20248_04
Continuation (5 of 10) of X20248 from base 400001 (Borrelia burgdorferi polynucleotide s
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

Query Match 14.3%; Score 31.4; DB 1; Length 110000;
Best Local Similarity 51.8%; Pred. No. 6.3;
Matches 71; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 9 GAATCTGTACAGAACCAACCAAGAGCTCCAAACCAACCAATCAAAATGCG 68
```

```
Db 21165 GAAACCTAGAACCTTATGATTAAGTTCATTATACATCATTAATAAATACTCAAAAATCTT 21224
Qy 69 AATCTCCAGACACATGGAGATCCGTGATTCGTGAGACTTTAGA 128
Db 21225 CATTATTATTAACGACTTTCTTTTATATAAAAAATCATTAACCAAAATAAATCATACCT 21284

Qy 129 ATGACAAACACAGCCACC 145
Db 21285 ATACACATAGCTATC 21301

RESULT 4
T35109
T35109 standard; cDNA; 524 BP.
ID T35109;
AC T35109;
DT 21-NOV-1996 (first entry)
DE Down-regulated senescence clone, SEND35.
KW Senescence related gene; expression; inhibition; acceleration; tomato;
KW lettuce; cabbage; banana; strawberry; wheat; maize; rice; rape; soybean;
KW delayed senescence; yield; protein content; quality; tolerance;
KW Lycopersicon esculentum.
OS W09507993-A1.
PN W09507993-A1.
PD 23-MAR-1995.
PF 13-SEP-1994; GB01990.
PR 13-SEP-1993; GB-018927.
PA (ZENE) ZENECA LTD.
PI Drake CR, Farrell A, Grierson D, Hosken SE, John I;
PI Schuch WW, Smart CM, Thomas H;
PI WPI: 95-131361/17.
DR DNA constructs which modify expression of senescence-related genes -
PT useful to accelerate or inhibit senescence in crop plants
PS Claim 14: Page 54; 78pp; English.
CC The sequences given in T35095-133 are senescence related genes and
CC fragments which were used in the construct of the invention which
CC modifies the expression of at least one senescence related gene.
CC Using these constructs senescence may be inhibited or accelerated
CC in plants including tomato, lettuce, cabbage, banana, strawberry,
CC wheat, maize, rice, rape or soybean. Delayed senescence may
CC indirectly prolong the life of the plant, increase yield, increase
CC protein content of fruits, improve quality of leafy vegetables,
CC improve tolerance to disease, drought or other stress. Increased
CC senescence may more rapidly break down unwanted plant material and
CC so avoid the use of desiccants on crops. This sequence represents the
CC senescence down-regulating clone, PSEND35, which is also known as 7256.
CC PSEND35 is a cDNA of approx. 0.7 kb. The mRNA encoded by PSEND35 is
CC expressed in green leaves of tomato plants, but at the onset of
CC senescence its expression is switched off. PSEND35 encodes a
CC protein of unknown function. PSEND35 is deposited as NCIMB 40580.
SQ Sequence 524 BP; 185 A; 90 C; 84 G; 163 T;

Query Match 14.1%; Score 31; DB 1; Length 524;
Best Local Similarity 53.8%; Pred. No. 1.9;
Matches 64; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 22 AACAAACCAACCAAGAGCTCCAAACCAAGCAAAATCAAAATCGAATCTCCGAACA 81
Db 21 AACAGAGACACAAAAGAAAACACAGAAATGCAAAATATACATCTAGAGGACGAA 80

Qy 82 CATGGGAAGTATCCGTGATTCGTAGGACTTTAAGAAACCACTCTATGACACACAG 140
Db 81 AATGGAAACAAACAATGACAATCATGATGTTGTCAACAAACAAACATATGCTATGAG 139

RESULT 5
Q24482/C
Q24482 standard; DNA; 1358 BP.
ID Q24482;
AC Q24482;
DT 04-NOV-1992 (first entry)
DE Stearoyl-ACP-desaturase (clone PDE57).
KW Delta-9-desaturase; stearoyl-ACP; oleoyl-ACP; lipid; biosynthesis; ds.
```



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 09:51:54 ; Search time 3268.17 Seconds  
(without alignments)  
217.005 Million cell updates/sec

Title: US-09-049-696-1

Perfect score: 223  
Sequence: 1 GAATTCACAGGAGATGATC.....ATCTGTGATCTGTTTGAG 223

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_p11: \*  
8: gb\_p12: \*  
9: gb\_p13: \*  
10: gb\_p14: \*  
11: gb\_p15: \*  
12: gb\_p16: \*  
13: gb\_p17: \*  
14: gb\_p18: \*  
15: gb\_p19: \*  
16: gb\_p20: \*  
17: gb\_p21: \*  
18: gb\_p22: \*  
19: gb\_p23: \*  
20: gb\_p24: \*  
21: gb\_p25: \*  
22: gb\_p26: \*  
23: gb\_p27: \*  
24: gb\_p28: \*  
25: gb\_p29: \*  
26: gb\_p30: \*  
27: gb\_p31: \*  
28: gb\_p32: \*  
29: gb\_p33: \*  
30: gb\_p34: \*  
31: gb\_p35: \*  
32: gb\_p36: \*  
33: gb\_p37: \*  
34: gb\_p38: \*  
35: gb\_p39: \*  
36: gb\_p40: \*  
37: gb\_p41: \*  
38: gb\_p42: \*  
39: gb\_p43: \*  
40: gb\_p44: \*  
41: gb\_p45: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	221	99.1	3311	11	AF039400	AF039400 Homo sapi

2	203	91.0	2826	42	AF127036	AF127036 Homo sapi
3	185	83.0	35278	11	AF039401	AF039401 Homo sapi
4	22	9.9	167880	11	HUT91322	HUT91322 Human Chrom
5	19	8.5	180388	9	HUMRETB1AS	L11910 Human retin
6	19	8.5	135928	10	HS434014	AL0223398 Homo sapi
7	19	8.5	190000	34	AC004479	AC004479 Homo sapi
8	19	8.5	180821	35	AC007115	AC007115 Homo sapi
9	19	8.5	55629	35	AC007593	AC007593 Drosophila
10	19	8.5	163362	35	AC007624	AC007624 Drosophila
11	19	8.5	40014	36	CELF10C1	U49831 Caenorhabdi
12	18	8.1	136170	11	AC004595	AC004595 Homo sapi
13	18	8.1	150332	11	AC004921	AC004921 Homo sapi
14	18	8.1	186986	11	AC006210	AC006210 Homo sapi
15	18	8.1	44851	12	AC004404	AC004404 Homo sapi
16	18	8.1	138187	12	MMMH461	AF027865 Mus muscu
17	18	8.1	7659	12	MUSMHCAC1	M11800 Mouse MHC c
18	18	8.1	526	12	MUSMHCAC1	M30159 Mouse MHC c
19	17	7.6	41625	1	SCH24	AL049826 Streptomy
20	17	7.6	12510	2	AE000532	AE000532 Helicobac
21	17	7.6	3317	3	AF001261	AF001261 Bos tauru
22	17	7.6	3288	3	AF001262	AF001262 Bos tauru
23	17	7.6	2820	3	AF001263	AF001263 Bos tauru
24	17	7.6	1265	3	AF001264	AF001264 Bos tauru
25	17	7.6	2984	3	BTU36445	U36445 Bos tauru
26	17	7.6	2155	5	A26494	A26494 Human IDUA
27	17	7.6	4475	5	A26497	A26497 Human IDUA
28	17	7.6	8141	6	BSF121LYS	X95646 Streptococc
29	17	7.6	71000	7	SPBP87	AL032684 S. pombe c
30	17	7.6	118335	8	AC006193	AC006193 Arabidops
31	17	7.6	5834	9	AB020631	AB020631 Homo sapi
32	17	7.6	156909	9	AB020633	AB020633 Homo sapi
33	17	7.6	100000	9	AB020875	AB020875 Homo sapi
34	17	7.6	161014	9	AP000088	AP000088 Homo sapi
35	17	7.6	100000	9	AP000141	AP000141 Homo sapi
36	17	7.6	301692	9	D87675	D87675 Homo sapien
37	17	7.6	133120	9	HS393P23	295400 Human DNA s
38	17	7.6	184974	9	HS435D1	286054 Homo sapien
39	17	7.6	114771	9	HS569D19	AL022334 Human DNA
40	17	7.6	107467	9	HS938B3	293023 Homo sapien
41	17	7.6	4480	9	HOMALIDNO2	M95740 Human alpha
42	17	7.6	2287	9	HOMPTPASE	M25393 Human prote
43	17	7.6	118456	10	HS191P20	AL034399 Human DNA
44	17	7.6	122557	10	HS229A8	286090 Human DNA s
45	17	7.6	183113	10	HS243E7	AL022323 Human DNA

## ALIGNMENTS

RESULT 1  
AF039400 3311 bp mRNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,  
DEFINITION complete cds.  
ACCESSION AF039400  
NID 94009457  
VERSION AF039400.1 GI:4009457  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM human.  
REFERENCE 1 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,R.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 3311)  
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.  
TITLE Direct Submission

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES Location/Qualifiers

source  
1..3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1..3311  
/gene="hCLCA1"  
/gene="hCLCA1"  
352..3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="pid:94009458"  
/db\_xref="gi:4009458"  
/translation="MGPKSSVFILLHLLEGALSNSLIQLNNNGEIVVAIDPNV  
EDFTLQIKMTQASLYLEATGKREYFNVAIIPETWTKADYVRPKLETYKNA  
DIVAYESTPRNDPEYTEMONGEGERHILPDPFAGKKAIEYPOGRAFEVMAH  
LWGVFDEYNDEKFIYLSNGRIQAVRSAGITGNNYKKGCGSCYTKRGTFFKVTGL  
YKGEFEVLOSROTOKASIMFAQVDSIVEFCIQNNKKEAPKQKNCMLRSTWPI  
RDSDEKRTTPTMTQPNPTFSLQIQRIYCLVLDSSGMATGNRLRLNQGQFL  
LQTVELGSWGMVFDSAAHVSQSLIQINSGSDRDTLAKRLPAASGGTISCGLRSA  
FVIRKRYTDSSEIVLTDGDNITSGCFNFKOSGAIITHVALGPSAODELSK  
DTYGLQYASDOVQNNGLIDAFGLSSGNVQSIOLESGLTLONSOMNGTVY  
MSTGKDTLFLITWTQPOILLMDPSGQGGFVVDKNTKMAVLIQIGIAKGTWY  
SLQASSQTLITVTSRASNATLPITVTSKTKDYSRFTYDNGRYSVYALGYN  
SVTALIESVNGKTVLELDNGAGADATKDDGVYSRFTYDNGRYSVYALGYN  
AARRVYIPQOAGALYIPGWIENDEIQNMPREINKDDVOHKVCSRTSGGSFVAS  
DVPNAPIDLPFGQITDKAEIHGSLINLTWTPAGDDYDHGTAHYIIRISTSLD  
LRKFNESLQVNTALIPKEANSEVEFLKPENTIFPENGDTLFIQAVKVDKSEI  
SNIRVSLFIPQTPETPSDPTSAPCPNHINSTIPGHIILKIMMKWIGELQSLIA

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 99.1%; Score 221; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 2, 9e-118;  
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AATCAAGGAGATGTACAGCAATGGGCCATTAGAGCTGTGTTCACTTAACTCT 62  
DB 330 AATCAAGGAGATGTACAGCAATGGGCCATTAGAGCTGTGTTCACTTAACTCT 389  
QY 63 TCACCTTTAGAAAGGGCCCTGAGTATTCATCTCATTCAGTGAACAATGGGTATGA 122  
DB 390 TCACCTTTAGAAAGGGCCCTGAGTATTCATCTCATTCAGTGAACAATGGGTATGA 449  
QY 123 AGGCAATGCTGTGCAATGAGCCCAATGTGCCAGAGATGAACAACATCAACAAAT 182  
DB 450 AGGCAATGCTGTGCAATGAGCCCAATGTGCCAGAGATGAACAACATCAACAAAT 509  
QY 183 AAGGACATGTGACCCAGGCAATCTGTATCTGTTGAAG 223  
DB 510 AAGGACATGTGACCCAGGCAATCTGTATCTGTTGAAG 550

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS AF127036  
DEFINITION Homo sapiens calcium-activated chloride channel protein 1 (CACCI)  
ACCESSION AF127036  
MRNA, complete cds.  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 2826)  
AUTHORS Agnel,M., Vernat,T. and Culouscou,J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride channel

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel,M. and Culouscou,J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Rueil-Malmaison 92500, France

FEATURES Location/Qualifiers

source  
1..2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1..2826  
/gene="CACCI"  
5..2749  
/gene="CACCI"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD25487.1"  
/db\_xref="pid:94585469"  
/db\_xref="gi:4585469"

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 91.0%; Score 203; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 9, 1e-108;  
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 AGCAATGGGCGCATTAAGAGTTGTGTCATCTTATCTTACCTTTAGAAAGGCG 80  
DB 1 AGCAATGGGCGCATTAAGAGTTGTGTCATCTTATCTTACCTTTAGAAAGGCG 60  
QY 81 CCTGAGTATCTCACTCAATTCAGTGAACAACAATGGCTATGAAGCAATGCTGTCAT 140  
DB 61 CCTGAGTATCTCACTCAATTCAGTGAACAACAATGGCTATGAAGCAATGCTGTCAT 120  
QY 141 CGACCCCAATGTGCCAGAGATGAACAACATCAACAATAAGGACATGGGACCCA 200  
DB 121 CGACCCCAATGTGCCAGAGATGAACAACATCAACAATAAGGACATGGGACCCA 180  
QY 201 GGCAATCTGTATCTGTTGAAG 223  
DB 181 GGCAATCTGTATCTGTTGAAG 203

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS AF039401  
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
complete cds.

ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber A.D., Eble R.C., Ji H.L., Schreuer K.D., Fuller C.M. and Pauli B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca<sup>2+</sup>-activated Cl<sup>-</sup> channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber A.D., Eble R.C. and Pauli B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source  
1.35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1.1617  
/gene="hCLCA1"  
1.33522  
/gene="hCLCA1"  
767.1265  
/repeat\_region="L1 transposable element"  
/rpt\_family="L1"  
/rpt\_type="dispersed"  
1587.1596  
/gene="hCLCA1"  
/note="putative"  
1618.1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC93429.1"  
/db\_xref="PID:94009460"  
/translation="MGPKSSVFLLIHLLEGALNSLIQLNNNGEIVAVIDPVP  
EDFTLLOIKKMTQATSLYFATGRKRFKPAVALIIPETWTKADYVPEKTYKNA  
DYLVAESTPPGNDPEPTTEOMNGCKGEKIHLPDITAKKLAETGPOKCAVHEAH  
LWGVFDEYNDEKFEYLSNGRIQAVKCSAGITGTNVKRCQGGSCYTKRCTFNKVTGL  
YKGECEFLVQSROTERKASIMFAOHVDSIVEFCTEONHNEAKNOKMCLNSTEVI  
RSEDEFKTPPTPTOPNPPTFSLQIGRIQVLDKSSMATGNRLNLNOGOLF  
LOTVELGSWGVATFDSAAHVSSELIQINSSGSDRTIARLPAASGGRSICGLRSA  
FTVIRKKYPTDGSSEIVLLTGDGMTSISGPKOSGALITHVAAGPNSAOLELEK  
MTGICLYASDOYVNNGLIDAFGALSSGCAVNSRSTQIESGLTIQNSOMKNGYIV  
DSTVGKDTLITWTQPOQILLMDPSGGKGGFVVDKTKNAVLIQIPGIADVWTKY  
SLQASSQTLITVTSRASNAFLPIVIVTSKTKNDTSKFPSPLVVYAINOGASPILRA  
SVTALIESVNGKTVLELDNGAGADATDDVYSRYFTYPTDNGRSYKVALGGVN  
AARRVYPOGSGALVPGWIENDEIQMNPPEINDVDYOHVOCFSRTSSGGSFYAS  
DVPNAPIPLPFPQGITDLNAELHGSGLNLMTWTAAGCDYDGTAKKTIIRISTILD

LDKPNESQVNTALLIPKEANSEVEFLKPNITFENCTDIFALQAVDKVLSKI  
SNIAKVSLEFPPTPEPTSPDETSAFCPNHINSTIPEHILKTKMKWIGELQLSIA  
"

6667..6807  
/gene="hCLCA1"  
/number=3  
6979..7126  
/gene="hCLCA1"  
/number=4  
9696..9801  
/gene="hCLCA1"  
/number=5  
15454..15631  
/gene="hCLCA1"  
/number=6  
18591..18809  
/gene="hCLCA1"  
/number=7  
19783..20010  
/gene="hCLCA1"  
/number=8  
22253..22427  
/gene="hCLCA1"  
/number=9  
24522..24628  
/gene="hCLCA1"  
/number=10  
26644..26859  
/gene="hCLCA1"  
/number=11  
27447..27708  
/gene="hCLCA1"  
/number=12  
28766..28936  
/gene="hCLCA1"  
/number=13  
31840..32079  
/gene="hCLCA1"  
/number=14  
32919..33522  
/gene="hCLCA1"  
/number=15

BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN

Query Match 83.0%; Score 185; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 2,8e-97;  
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AATCAGAGGAGATGTACAGCAATGGGCGCATTAAGAGTCTGTGTCATCTGATCT 62  
|||||  
DB 2200 AATCAGAGGAGATGTACAGCAATGGGCGCATTAAGAGTCTGTGTCATCTGATCT 2259  
|||||

QY 63 TCACCTTCTAGAGGGGCCCTGAGTAATTCACATTCAGCTGCAACAAATGGCTATCA 122  
|||||  
DB 2260 TCACCTTCTAGAGGGGCCCTGAGTAATTCACATTCAGCTGCAACAAATGGCTATCA 2319  
|||||

QY 123 AGGCAATGCTGTCGAATGAGCAATGATGCGAGAAATGAAACACTGATTCACAAAT 182  
|||||  
DB 2320 AGGCAATGCTGTCGAATGAGCAATGATGCGAGAAATGAAACACTGATTCACAAAT 2379  
|||||

QY 183 AAAG 187  
|||||  
DB 2380 AAAG 2384

RESULT 4  
HUU91323  
LOCUS HUU91323 167880 bp DNA  
DEFINITION Human Chromosome 16 BAC clone C198785K-A-972D3, complete sequence.  
ACCESSION U91323  
NID 93582311

VERSION D91323.1 GI:3582311  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 167880)  
 AUTHORS Adams M.D., Loftus B.J., Phillips C.A., Zhou L., Brandon R. and Venter J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-MAR-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
 REFERENCE 2 (bases 1 to 167880)  
 AUTHORS Adams M.D., Loftus B.J., Zhou L., La Bombard M., Kim U.J. and Venter J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
 REFERENCE 3 (bases 1 to 167880)  
 AUTHORS Adams M.D., Loftus B.J., Zhou L., La Bombard M., Kim U.J. and Venter J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
 REFERENCE 4 (bases 1 to 167880)  
 AUTHORS Adams M.D. and Loftus B.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-SEP-1998) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
 COMMENT On Sep 11, 1998 this sequence version replaced gi:2335056. BAC clone C1987SK-972D3 is located in band 16p13.11 of chromosome 16. Genes were identified by a combination of five methods: XGRL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (available by anonymous ftp from colin@u.washington.edu), GENSCAN (available using the e-mail server at genscan@genomic.stanford.edu), searches of the EST database at TIGR (http://www.tigr.org/tcd/hcd/hcd.html) and searches against a peptide database. Repeats were identified using RepeatMasker (sm1t, A.F.A. and Green, P. unpublished, http://ftp.genome.washington.edu/rm/RepeatMasker.html).  
 FEATURES  
 source location/Qualifiers  
 1.167880  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /map="16p13.11"  
 /clone="A-972D3"  
 68302..68901  
 /gene="A-972D3.1"  
 /pseudogene  
 <162350..>162696  
 /gene="A-972D3.2"  
 <162350..>162696  
 /gene="A-972D3.2"  
 162350..>162696  
 /gene="A-972D3.2"  
 /codon\_start=1  
 /product="Myosin heavy chain (MYH11) (3'partial)."  
 /protein\_id="AAC35212.1"  
 /db\_xref="pid:g1871206"  
 /translation="MAKQGLSDDEKFLVDKFNISPVADAMAKRILVWVSEKOG  
 FEASIKKEKGEDEVVELVBNKKTIVGKDDIOKMPKPFSEYEDMAELTCLNEASVL  
 HNLRRYFSLIV"  
 BASE COUNT 46429 a 38556 c 38892 g 43951 t 52 others  
 ORIGIN  
 Query Match 9.98; Score 22; DB 11; Length 167880;  
 Best Local Similarity 100.0%; Pred. No. 0.027;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CDS 164 AACACGCTATTCACACAAATAAA 185

Db 25031 AACACGCTATTCACACAAATAAA 25052  
 RESULT 5  
 LOCUS HUMRETBLAS/180388 bp DNA PRI 23-NOV-1994  
 DEFINITION Human retinoblastoma susceptibility gene exons 1-27, complete cds.  
 ACCESSION L11910  
 MID G292420  
 VERSION L11910.1 GI:292420  
 KEYWORDS nuclear protein; recessive oncogene; retinoblastoma gene; retinoblastoma protein; retinoblastoma susceptibility; tumor suppressor gene.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Friend S.H., Bernards R., Rogel J.S., Weinberg R.A., Rapaport J.M., Albert D.M. and Dryja T.P.  
 TITLE A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma  
 JOURNAL Nature 323 (6089), 643-646 (1986)  
 MEDLINE 87039336  
 REFERENCE 2 (sites)  
 AUTHORS Friend S.H., Horowitz J.M., Gerber M.R., Wang X.F., Bogenmann E., Li F.P. and Weinberg R.A.  
 TITLE Deletions of a DNA sequence in retinoblastomas and mesenchymal tumors: organization of the sequence and its encoded protein  
 JOURNAL [published erratum appears in Proc Natl Acad Sci U S A 1988 Apr;85(7):2234]  
 MEDLINE 88097427  
 REFERENCE 3 (sites)  
 AUTHORS Lee W.H., Bookstein R., Hong F., Young L.J., Shew J.Y. and Lee E.Y.  
 TITLE Human retinoblastoma susceptibility gene: cloning, identification, and sequence  
 JOURNAL Science 235 (4794), 1394-1399 (1987)  
 MEDLINE 87149066  
 REFERENCE 4 (sites)  
 AUTHORS McGee T.L., Yandell D.W. and Dryja T.P.  
 TITLE Structure and partial genomic sequence of the human retinoblastoma susceptibility gene  
 JOURNAL Gene 80 (1), 119-128 (1989)  
 MEDLINE 90006771  
 REFERENCE 5 (sites)  
 AUTHORS Sakai T., Ohtani N., McGee T.L., Robbins P.D. and Dryja T.P.  
 TITLE Oncogenic germ-line mutations in Spl and Atf sites in the human retinoblastoma gene  
 JOURNAL Nature 353 (6339), 83-86 (1991)  
 MEDLINE 91351319  
 REFERENCE 6 (bases 1 to 180388)  
 AUTHORS Touchida J., McGee T.L., Paterson J.C., Eagle J.R., Tucker S., Yandell D.W. and Dryja T.P.  
 TITLE Complete genomic sequence of the human retinoblastoma susceptibility gene  
 JOURNAL Genomics 17 (3), 535-543 (1993)  
 MEDLINE 94063891  
 FEATURES  
 source location/Qualifiers  
 1..180388  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 185..1863  
 /bound\_moiety="Spi"  
 /evidence="experimental"  
 1866..1872  
 /bound\_moiety="ATP"  
 /evidence="experimental"  
 <2060..>2196  
 /number=1  
 /join(2060..2196,5424..5550,39446..39561,41926..42045,  
 44668..44706,45799..45866,56853..56963,59651..59793,

61730.. .61807,64330.. .64439,65364.. .65441,70242.. .70329,  
73753.. .73869,76430.. .76486,76889.. .76920,77001.. .77077,  
78083.. .78279,149998.. .150116,153208.. .153253,156693.. .156838  
160730.. .160834,161997.. .162110,162204.. .162367,  
170372.. .170402,173707.. .173849,174361.. .174410,  
177005.. .177078)  
/codon\_start=1  
/product="retinoblastoma susceptibility protein"  
/protein\_id="AAAS3483.1"  
/db\_xref="GI:292421"  
/translation="MPPTPRKRTATATAA AAPAPPPPEEDPEDDSCGDELPL  
VRLPEEPEEDPFALCOKLKIPDHVERAMLMMEKVSSVDGLGGYIOCKREIMAC  
IFIAAVLDDEDSFTFLQKNIELSVKRFNLKEIDLTSTKYNDAMSRLIKRYDYFA  
LFSULERCELIITLOPSSSISTEINSALVKYSWTFELAKGEVLQMEDLVISFOL  
MLCVADIVFKISPPMLAKEPKTAVIPINSGSPTRPGONRSARIAQLQENDRIILEV  
LCRKHENIDEVKNVYKFNLPFNNSIGLVTSNGLPEVENLSKRVEIYLKNNDLAR  
LELDHDTLOTDSIDSEFOTRTPSKSLUDEEVNIPIHPTEVRVNTTIOOLMILNSA  
SDOSENLIQSFNNCTVPNKRESILTKRVKDGYIEKKFAAVAGOGCVELIGSORRYKTG  
RLTYRWMSMKASEERNLSTIONSKLINDIIFHSLLACLAEVYMAITSRSIQNIDS  
GTDSLFPWILNWKAFDFYKYESTIAEGNLTIRMILHLRCCHRTMESLAWLSD  
SPDLDLKOSDRGPDHLESACPMLPIQNHTADAYTLSEVRSPKKGSTYRWS  
TANAEQTASFOTOKLFKSTLSLIFYKVTYRAYLRNLTKSEPELEHIIWT  
LEOFLOTNEYELMRDLDOIOMCSMYGICKVNIDLKFIITYAKDLPHAVOEYFK  
PVLLKEEYSDSIYFVNSVMORLKTNILOYASTRPETLSPIHIIRSPYQPCPSBLR  
IPGNIIYTSPISKPYKISEGLPTPTTKTTPRSRLIVSIGESFGTSEKFQAINOVXSD  
RVLRSLAESGSPPKPLKRLNRDIEGSDEAGSKHLPEESKFOOKLAEMTSTRIMKO  
KMNSMDTSMEEK"  
2197.. .5423  
/number=1  
2302  
/standard\_name="Restriction Fragment Length Polymorphism"  
/note="BamBI RFLP"  
5424.. .5550  
/number=2  
5551.. .39445  
/number=2  
14868.. .14914  
/standard\_name="microsatellite RB12 (AFM058xd6)"  
39446.. .39561  
/number=3  
39562.. .41925  
/number=3  
41926.. .42045  
/number=4  
42046.. .44667  
/number=4  
43218.. .43257  
/standard\_name="microsatellite RB14"  
44668.. .44706  
/number=5  
44707.. .45798  
/number=5  
45799.. .45866  
/number=6  
45867.. .56852  
/number=6  
56853.. .56963  
/number=7  
56964.. .59650  
/number=7  
59651.. .59793  
/number=8  
59794.. .61729  
/number=8  
61730.. .61807  
/number=9  
61808.. .64329  
/number=9  
64330.. .64439  
/number=10  
64440.. .65363  
/number=10

```

      exon          65364. .65441
                    /number=11
      Intron        65442. .70241
                    /number=11
      exon          70242. .70329
                    /number=12
      Intron        70330. .73752
                    /number=12
      exon          73753. .73869
                    /number=13
      Intron        73870. .76429
                    /number=13
      exon          76430. .76486
                    /number=14
      Intron        76487. .76888
                    /number=14
      exon          76889. .76920
                    /number=15
      Intron        76921. .77000
                    /number=15
      exon          77001. .77077
                    /number=16
      Intron        77078. .78082
                    /number=16
      exon          78083. .78279
                    /number=17
      Intron        78280. .14997
                    /number=17
      variation     99426
                    /standard_name="Restriction Fragment Length Polymorphism"
                    /note="XbaI RFLP"
      satellite     102921. .102959
                    /standard_name="microsatellite RB117"
                    123912. .125501
      repeat_region /standard_name="Variable Number Tandem Repeats (VNTR)"
                    /rpt_type=tandem
      exon          149998. .150116
                    /number=18
      Intron        150117. .153207
                    /number=18
      exon          153208. .153353
                    /number=19
      Intron        153354. .156692
                    /number=19
      exon          156693. .156838
                    /number=20
      Intron        156839. .160729
                    /number=20
      satellite     156895. .156976
                    /standard_name="microsatellite RB1.20"
      exon          160730. .160834
                    /number=21
      Intron        160835. .161996
                    /number=21
      exon          161997. .162110
                    /number=22
      Intron        162111. .162203
                    /number=22
      exon          162204. .162367
                    /number=23

Query Match      8.5%; Score 19; DB 9; Length 180388;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 GACCCAGGCATCTCTGTAT 213
|||||
DB 154951 GACCCAGGCATCTCTGTAT 154933

RESULT 6
HS434014 135928 bp DNA PRI 10-JUN-1999
LOCUS
```





/gene="HSD11B1"  
/note="clone 447D17; GAA in this entry; substitution"  
/replace="gga"  
12557..12559  
/gene="HSD11B1"  
/note="clone 447D17; ATA in this entry; substitution"  
/replace="aga"  
14614..14616  
/note="clone 447D17; ATC in this entry; substitution"  
/replace="acc"  
14736..15026  
/note="AluSC repeat: matches 2..292 of consensus"  
16717..16750  
/note="17 copies of 2 mer 82 & conserved"  
16745..16748  
/note="clone 447D17; GTGT in this entry; insertion"  
/replace="gt"  
16753..17060  
/note="AluIO repeat: matches 300..1 of consensus"  
17303..17321  
/note="11MD2 repeat: matches 459..673 of consensus"  
17472..17474  
/note="clone 447D17; TGT in this entry; substitution"  
/replace="tat"  
17684..18138  
/note="MTRIC repeat: matches 1..464 of consensus"  
17726..17728  
/note="clone 447D17; CTC in this entry; substitution"  
/replace="cgc"  
17838..17840  
/note="clone 447D17; CAT in this entry; substitution"  
/replace="cgt"  
18278..18280  
/note="clone 447D17; ACC in this entry; substitution"  
/replace="atc"  
18662..18743  
/note="11P83 repeat: matches 70..150 of consensus"  
18744..19043  
/note="AluSX repeat: matches 2..302 of consensus"  
18781..18783  
/note="clone 447D17; TCG in this entry; substitution"  
/replace="ttg"  
19009..19011  
/note="clone 447D17; ATG in this entry; substitution"  
/replace="acg"  
19044..19830  
/note="11P81 repeat: matches 132..902 of consensus"  
19444..19451  
/note="clone 447D17; TACACACA in this entry; insertion"  
/replace="ta"  
19911..20210  
/note="AluIO repeat: matches 302..1 of consensus"  
20147..20149

Query Match 8.5%; Score 19; DB 10; Length 135928;  
Best Local Similarity 100.0%; Fred. No. 1.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 165 AACACTCATTACACAATA 183  
|||||  
DB 81633 AACACTCATTACACAATA 81651

RESULT 7  
AC004479/c AC004479 190000 bp DNA HTG 02-MAR-1999  
LOCUS Home sapiens chromosome 4, WORKING DRAFT SEQUENCE, 50 unordered  
DEFINITION  
pieces.  
AC004479 AC004479  
NID 94314347  
VERSION AC004479.2 GI:4314347  
KEYWORDS HTG: HTGS\_PHASE1.  
SOURCE human.

ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1. (bases 1 to 190000)  
JOURNAL Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.  
REFERENCE Direct Submission  
AUTHORS Unpublished  
JOURNAL 2 (bases 1 to 190000)  
REFERENCE Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.  
JOURNAL Direct Submission  
COMMENT Submitted (27-MAR-1998) Department of Genetics, Stanford Human  
Genome Center, 885 California Avenue, Palo Alto, CA 94304, USA  
On Mar 2, 1999 this sequence version replaced gi:4225908.  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 50 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 1049 1048: contig of 1048 bp in length  
\* 1177 1176: gap of unknown length  
\* 2584 2583: contig of 1407 bp in length  
\* 2712 2711: gap of unknown length  
\* 4171 4170: contig of 1459 bp in length  
\* 4299 4298: gap of unknown length  
\* 5775 5774: contig of 1476 bp in length  
\* 5903 5902: gap of unknown length  
\* 7025 7024: contig of 1122 bp in length  
\* 7153 7152: gap of unknown length  
\* 8363 8362: contig of 1083 bp in length  
\* 8364 8363: gap of unknown length  
\* 9589 9588: contig of 1225 bp in length  
\* 9717 9716: gap of unknown length  
\* 11307 11306: contig of 1490 bp in length  
\* 11335 11334: gap of unknown length  
\* 12505 12504: contig of 1170 bp in length  
\* 12633 12632: gap of unknown length  
\* 14579 14578: contig of 1946 bp in length  
\* 14707 14706: gap of unknown length  
\* 15907 15906: contig of 1200 bp in length  
\* 16035 16034: gap of unknown length  
\* 18141 18141: contig of 2107 bp in length  
\* 18142 18269: gap of unknown length  
\* 18270 19573: contig of 1304 bp in length  
\* 19574 19701: gap of unknown length  
\* 19702 21240: contig of 1539 bp in length  
\* 21241 21368: gap of unknown length  
\* 21369 21366: contig of 1998 bp in length  
\* 23367 23494: gap of unknown length  
\* 23495 24828: contig of 1334 bp in length  
\* 24829 24956: gap of unknown length  
\* 24957 26280: contig of 1324 bp in length  
\* 26281 26408: gap of unknown length  
\* 26409 27622: contig of 1214 bp in length  
\* 27623 27750: gap of unknown length  
\* 27751 29200: contig of 1450 bp in length  
\* 29201 29328: gap of unknown length  
\* 29329 31183: contig of 1855 bp in length  
\* 31184 31311: gap of unknown length  
\* 31312 32868: contig of 1657 bp in length  
\* 32869 33096: gap of unknown length  
\* 33097 34821: contig of 1725 bp in length  
\* 34822 34949: gap of unknown length  
\* 34950 36652: contig of 1703 bp in length  
\* 36653 36780: gap of unknown length  
\* 36781 38395: contig of 1615 bp in length  
\* 38396 38523: gap of unknown length  
\* 38524 39706: contig of 1183 bp in length  
\* 39707 39834: gap of unknown length  
\* 39835 41692: contig of 1858 bp in length  
\* 41693 41820: gap of unknown length

\* 41821 44132: contig of 2312 bp in length  
\* 44133 44260: gap of unknown length  
\* 44261 45715: contig of 1455 bp in length  
\* 45716 45843: gap of unknown length  
\* 45844 48070: contig of 2227 bp in length  
\* 48071 48198: gap of unknown length  
\* 48199 49385: contig of 1187 bp in length  
\* 49386 49513: gap of unknown length  
\* 49514 52572: contig of 3059 bp in length  
\* 52573 52700: gap of unknown length  
\* 52701 54823: contig of 2123 bp in length  
\* 54824 54951: gap of unknown length  
\* 54952 57135: contig of 2184 bp in length  
\* 57136 57263: gap of unknown length  
\* 57264 59168: contig of 1905 bp in length  
\* 59169 59296: gap of unknown length  
\* 59297 61788: contig of 2492 bp in length  
\* 61789 61916: gap of unknown length  
\* 61917 65782: contig of 3866 bp in length  
\* 65783 65910: gap of unknown length  
\* 65911 68324: contig of 2414 bp in length  
\* 68325 68452: gap of unknown length  
\* 68453 71054: contig of 2602 bp in length  
\* 71055 71182: gap of unknown length  
\* 71183 73489: contig of 2307 bp in length  
\* 73490 73617: gap of unknown length  
\* 73618 76516: contig of 2899 bp in length  
\* 76517 76644: gap of unknown length  
\* 76645 81029: contig of 4385 bp in length  
\* 81030 81157: gap of unknown length  
\* 81158 87299: contig of 6142 bp in length  
\* 87300 87427: gap of unknown length  
\* 87428 95665: contig of 8238 bp in length  
\* 95666 95793: gap of unknown length  
\* 95794 104477: contig of 8684 bp in length  
\* 104478 104605: gap of unknown length  
\* 104606 119197: contig of 14592 bp in length  
\* 119198 119325: gap of unknown length  
\* 119326 125177: contig of 5852 bp in length  
\* 125178 125305: gap of unknown length  
\* 125306 138807: contig of 13502 bp in length  
\* 138808 138934: gap of unknown length  
\* 138935 149110: contig of 10176 bp in length  
\* 149111 149237: gap of unknown length  
\* 149238 164669: contig of 15432 bp in length  
\* 164670 164797: gap of unknown length  
\* 164797 190000: contig of 25204 bp in length.

FEATURES  
source  
1.190000 Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"

BASE COUNT 45019 a 47468 c 47837 g 43356 t 6320 others  
ORIGIN

Query Match 8.5%: Score 19; DB 34; Length 190000;  
Best local Similarity 100.0%: Pred. No. 1.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 AACACATCAACAATA 183  
|||||  
Db 149929 AACACATCAACAATA 149911

RESULT 8  
AC007115 180821 bp DNA HTG 27-MAY-1999  
LOCUS Homo sapiens chromosome 12 clone 91705, WORKING DRAFT SEQUENCE, 1n  
DEFINITION Ordered pieces.  
AC007115  
AC007115 54895146  
NID  
VERSION AC007115.1 GI:4895146

KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 180821)  
AUTHORS Montgomery, K.T., Lau, S.T. and Kucherlapati, R.  
TITLE High Throughput Sequencing of Human Chromosome 12  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 180821)  
AUTHORS Montgomery, K.T., Lau, S.T. and Kucherlapati, R.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAY-1999) Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461, USA  
COMMENT \* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1.180821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="91705"  
/clone\_1lb="RPG111"

BASE COUNT 59926 a 31181 c 31322 g 58392 t  
ORIGIN

Query Match 8.5%: Score 19; DB 35; Length 180821;  
Best local Similarity 100.0%: Pred. No. 1.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GTGCCAGAGATGAACAC 169  
|||||  
Db 22328 GTGCCAGAGATGAACAC 22346

RESULT 9  
AC007593/c 55629 bp DNA HTG 24-MAY-1999  
LOCUS Drosophila melanogaster chromosome 2 clone BACR01C10 (p620) RPT-98  
DEFINITION 01.C.10 map 42E-43A strain y2; cn bw sp, WORKING DRAFT SEQUENCE, 29  
unordered pieces.  
AC007593  
94885670  
AC007593.3 GI:4885670  
NID HTG: HTGS\_PHASE2.  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Drosophila melanogaster  
fruit fly.  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 55629)  
AUTHORS Ceoliker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,  
Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karia, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshireli, A.R., Moshireli, M., Nixon, K., Pacled, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequerra, A., Sethi, H., Snit, E.,  
Svilaras, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.T. and  
Rubin, G.M.  
TITLE Sequencing of Drosophila melanogaster  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 55629)  
AUTHORS Ceoliker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,  
Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karia, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshireli, A.R., Moshireli, M., Nixon, K., Pacled, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequerra, A., Sethi, H., Snit, E.,

TITLE  
JOURNAL  
COMMENT

Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.  
 Direct Submission  
 Submitted (20-May-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On May 24, 1999 this sequence version replaced gi:4883580.  
 For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet the following cutoffs: length >= 400 bases, phrap computed error rate <= 1/10.  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
 1 738: contig of 738 bp in length  
 739 818: gap of unknown length  
 819 1565: contig of 747 bp in length  
 1566 1645: gap of unknown length  
 1646 2488: contig of 843 bp in length  
 2489 2568: gap of unknown length  
 2569 3255: contig of 687 bp in length  
 3256 3335: gap of unknown length  
 3336 3894: contig of 559 bp in length  
 3895 3974: gap of unknown length  
 3975 5073: contig of 1099 bp in length  
 5074 5153: gap of unknown length  
 5154 6281: contig of 1128 bp in length  
 6282 6361: gap of unknown length  
 6362 7274: contig of 913 bp in length  
 7275 7354: gap of unknown length  
 7355 8930: contig of 1576 bp in length  
 8931 9010: gap of unknown length  
 9011 9924: contig of 914 bp in length  
 9925 10004: gap of unknown length  
 10005 11258: contig of 1254 bp in length  
 11259 11338: gap of unknown length  
 11339 12185: contig of 847 bp in length  
 12186 12265: gap of unknown length  
 12266 13482: contig of 1217 bp in length  
 13483 13562: gap of unknown length  
 13563 14741: contig of 1179 bp in length  
 14742 14821: gap of unknown length  
 14822 15969: contig of 1148 bp in length  
 15970 16049: gap of unknown length  
 16050 17736: contig of 1687 bp in length  
 17737 17816: gap of unknown length  
 17817 18892: contig of 1076 bp in length  
 18893 18972: gap of unknown length  
 18973 20387: contig of 1415 bp in length  
 20388 20467: gap of unknown length  
 20468 21779: contig of 1312 bp in length  
 21780 21859: gap of unknown length  
 21860 23391: contig of 1532 bp in length  
 23392 23471: gap of unknown length  
 23472 25449: contig of 1978 bp in length  
 25450 25529: gap of unknown length  
 25530 29096: contig of 3567 bp in length  
 29097 29176: gap of unknown length  
 29177 31225: contig of 2049 bp in length  
 31226 31305: gap of unknown length  
 31306 34234: contig of 2929 bp in length  
 34235 34314: gap of unknown length  
 34315 37164: contig of 2850 bp in length  
 37165 37244: gap of unknown length  
 37245 41254: contig of 4010 bp in length  
 41255 41334: gap of unknown length  
 41335 45340: contig of 4006 bp in length

FEATURES  
 source  
 1. 55629  
 /organism="Drosophila melanogaster"  
 /strain="y2; cn bw sp"  
 /db\_xref="taxon:7227"  
 /clone="BACR01C10 (D620) RPCI-98 01.C.10"  
 /chromosome="2"  
 /clone.lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"  
 /map="42E-43A"  
 /map="42E-43A"  
 BASE COUNT 14037 a 12352 c 12473 g 14526 t 2241 others  
 ORIGIN  
 Query Match 8.5%; Score 19; DB 35; Length 55629;  
 Best Local Similarity 100.0%; Pred. No. 1.5;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 161 ATGAACACTCATTTCACA 179  
 |||||  
 Db 28872 ATGAACACTCATTTCACA 28854  
 RESULT 10  
 AC007624  
 LOCUS  
 DEFINITION  
 AC007624 163362 bp DNA HTG 25-MAY-1999  
 Drosophila melanogaster chromosome 2 clone BACR1015 (D621) RPCI-98 10.F.15 map 42E-43A strain y2; cn bw sp, WORKING DRAFT SEQUENCE, 20 unnumbered pieces.  
 AC007624  
 NID 94887256  
 VERSION AC007624.3 GI:4887256  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE  
 ORGANISM  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 163362)  
 REFERENCE  
 AUTHORS  
 Celniker, S.E., Agdayani, A., Arcalata, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotar, M.A., Mazda, P., Moshirefi, A.R., Moshirefi, M., Nixon, K., Pacled, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Sitr, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.  
 Sequencing of Drosophila melanogaster  
 Unpublished  
 2 (bases 1 to 163362)  
 REFERENCE  
 JOURNAL  
 AUTHORS  
 Celniker, S.E., Agdayani, A., Arcalata, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotar, M.A., Mazda, P., Moshirefi, A.R., Moshirefi, M., Nixon, K., Pacled, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Sitr, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.  
 Direct Submission  
 Submitted (21-May-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On May 25, 1999 this sequence version replaced gi:4885669.  
 For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet

the following cutoffs: length >= 400 bases, phrap computed error rate <= 1/10.

\* NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 846: contig of 846 bp in length
847 926: gap of unknown length
927 1944: contig of 1018 bp in length
1945 2024: gap of unknown length
2025 3619: contig of 1595 bp in length
3620 3699: gap of unknown length
3700 5574: contig of 1875 bp in length
5575 5654: gap of unknown length
5655 7129: contig of 1475 bp in length
7130 9028: gap of unknown length
9029 9108: contig of 1819 bp in length
9109 11392: gap of unknown length
11393 11472: gap of unknown length
11473 15071: contig of 3599 bp in length
15072 15151: gap of unknown length
15152 17100: contig of 1949 bp in length
17101 20327: gap of unknown length
17181 20407: contig of 3147 bp in length
20328 20407: gap of unknown length
20408 26429: contig of 6022 bp in length
26430 26509: gap of unknown length
26510 32656: contig of 6147 bp in length
32657 32736: gap of unknown length
32737 38186: contig of 5450 bp in length
38187 38266: gap of unknown length
38267 47759: contig of 9493 bp in length
47760 47839: gap of unknown length
47840 54358: contig of 6519 bp in length
54359 54438: gap of unknown length
54439 62971: contig of 8533 bp in length
62972 63051: gap of unknown length
63052 76828: contig of 13777 bp in length
76829 76908: gap of unknown length
76909 104731: contig of 27823 bp in length
104732 104811: gap of unknown length
104812 128109: contig of 23298 bp in length
128110 128189: gap of unknown length
128190 163362: contig of 35173 bp in length.
Location/Qualifiers
1.163362
/organism="Drosophila melanogaster"
/strain="y2: cn bw sp"
/db_xref="taxon:7227"
/clone="BACR10F15 (D621) RPCI-98 10.F.15"
/chromosome="2"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC3.6"
/map="42E-43A"

```

BASE COUNT 44309 a 36102 c 35977 g 45454 t 1520 others

Query Match 8.5%: Score 19; DB 35; Length 163362;  
Best Local Similarity 100.0%: Pred No. 1.5; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 ATGAACACTCATTCACAA 179  
|||||  
DB 92274 ATGAACACTCATTCACAA 92292

RESULT 11  
CELLF10C1  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Caenorhabditis elegans strain-Bristol N2.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
Rhabditina; Rhabditidae; Rhabditidae; Pelodermatidae; Caenorhabditis.

REFERENCE  
AUTHORS  
1 (bases 1 to 40014)  
Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,  
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,  
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,  
Fullon, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,  
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Lalster, N.,  
Latter, P., Lightning, J., Lloyd, C., McKurray, A., Mortimore, B.,  
O'Callaghan, M., Parsons, J., Percy, C., Riken, L., Roopra, A.,  
Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,  
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,  
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,  
Wilkinson-Spratt, J. and Wohldman, P.  
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans  
Nature 368 (6466), 32-38 (1994)  
94150718  
REFERENCE  
AUTHORS  
TITLE  
The sequence of C. elegans cosmid F10C1  
Unpublished (1996)  
3 (bases 1 to 40014)  
Waterston, R.  
DIRECT SUBMISSION  
Submitted (20-FEB-1996) Robert Waterston  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
e-mail: tw@genome.wustl.edu and jessesanger.ac.uk  
NEIGHBORING COSMID INFORMATION:  
The 5' cosmid is C18A3, 199 bp overlap; 3' cosmid is K05F1, 200 bp  
overlap. Actual start of this cosmid is at base position 196 of  
CELLF10C1; actual end is at 6257 of CELK05F1

NOTES:  
Coding sequences below are predicted from computer analysis, using  
the program GeneFinder (P. Green and L. Hillier, ms in  
preparation).  
Location/Qualifiers

FEATURES  
source  
1.40014  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/map="11"  
6568..112916  
/gene="F10C1.7b"  
6568..10346  
/gene="F10C1.7a"  
join(6568..6631,7610..7803,8090..8211,8257..8616,  
8670..9089,9185..9235,9823..9970,10024..10131,  
10182..10346)  
/gene="F10C1.7a"  
/note="coded for by C. elegans cDNA CEMSC13F; coded for by  
C. elegans cDNA cm0198; coded for by C. elegans cDNA  
cm065; coded for by C. elegans cDNA cm14910; coded for by  
C. elegans cDNA yk3d10.3; coded for by C. elegans cDNA  
yk494.5; coded for by C. elegans cDNA yk494.3; similar to  
intermediate filament protein; alternative splicing to  
F10C1.7b"  
/codon\_start=1

/protein\_id="AA093405.1"  
/db\_xref="pid:g1203948"  
/db\_xref="gi:1203948"  
/translation="MSAVSYSMHRTTTTSSSSHGVSAGHAEEFVASAREKQEMQ  
OLNSLEYVYISRLQEDLNKELVELDTLRSLNDIGOIATKFNDSLVKRRISE  
AHSGLTVEYISRLQEDLNKELVELDTLRSLNDIGOIATKFNDSLVKRRISE  
AALDEEKRIVAEODLYLOLAARKEIDALVDRRIQAEEDDKITELFGRTHSO  
ETTELRTLAQAPADTRREFKRNELALREIKAEKDTITQTRVLETFIPOSKISAVE  
SSIVSKNEAAVROEIKRMNESITTLAKLSLEARNASALEREANTQIQIGEOQRA  
YESELKRDNALRFREDCOTLIQALQALNTKOTIDTEIARIVLSEEGRTFHV  
OGVVAOQETRLVVEODHMDGSEVOIRSSFRKAKGNVSIVEDCDPOGKYLIENTS  
GSVAEDVNFETIRIVYDGOAFVFRPLPSHLVIOGHKIKYGRNNGELINLPDSIVME  
SHASWGCGRAETPIYNSHGIEKASHIILIDRFQLLHPSN"  
complement(1384, .14789)  
/gene="F10c1.3"  
/complement(join(1384, .13503,14616, .14789))  
/gene="F10c1.3"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AA093406.1"  
/db\_xref="pid:g1203949"  
/db\_xref="gi:1203949"  
/translation="MLYMLAVFDFSMVLRNRIHALVTCIOCTLISEHGLTAIEDRDT  
GTRKSTFRVSSDALATVLPAYAIPEGKKGKTKGKKNKSYEDDLAAGDK"  
22010, .28059  
/gene="F10c1.2b"  
join(22010, .22142,25467, .25717,26015, .26136,26182, .26970,  
27081, .27282,27330, .27437,27895, .28059)  
/gene="F10c1.2b"  
/note="coded for by C. elegans CDNA CEES081f; coded for by  
C. elegans CDNA CESA13f; coded for by C. elegans CDNA  
yk36.3; coded for by C. elegans CDNA yk1267.3; coded for  
by C. elegans CDNA yk1267.5; coded for by C. elegans CDNA  
yk37c3.5; similar to intermediate filament protein  
(alternative splicing to F10c1.2)"  
/codon\_start=1  
/protein\_id="AA093407.1"  
/db\_xref="pid:g1203950"  
/db\_xref="gi:1203950"  
/translation="MSHKESSSEYEMQRYSTIOPRTAVRSOSRGNTVSGNGAGSG  
GRVLKMTVMGSAIVGIGSPALSAANAKSPFEAKDEKKTLOGLDRLGNIDRVKRL  
EONRKLIVADIDELRGKWKDTSPIKIVSSLSIARDDIDDAARAKAEVYKVAHR  
DDLAERSRYEDIOORRESDEKTSQRTNATADQSEVEMRAPRKQUTDEKRYTAD  
NSRIWELOKARSDIDETIGRIDQONOVOLMELEFLURVHDEVEVIELALAQAP  
ADTRFEFFKNELALAIIDIKEDYDIIAKQKQDMESWYLVKVSOGSARNANMESTYO  
RDEVKRMNDNIGDLRGKLDLENKNSLIEKEVONVLTQIDDOQRYEALNDPRATLR  
RMRECOITLVAELQALDITKQMLDAEIIAYIRKMLEGETRVLQOMVPOAVKTSILOO  
QENDSTRSVNGEVSTKTTFQSAKGNVTISECDPENGKFIENSHRNKNDVNGEHI

RRKLDGREIYYSIPANVYTRGKNTLTIYADOGGINNPPESLYFDGENTWIGANV  
TSLVKNDEERATHHTOKTIOGQ.  
25329, .28059  
/gene="F10c1.2a"  
join(25329, .25368,25467, .25717,26015, .26136,26182, .26970,  
27081, .27282,27330, .27437,27895, .28059)  
/gene="F10c1.2a"  
/note="coded for by C. elegans CDNA CEES081f; coded for by  
C. elegans CDNA CESA13f; coded for by C. elegans CDNA  
yk13c11.3; coded for by C. elegans CDNA yk36.3; coded for  
by C. elegans CDNA yk13c11.5; coded for by C. elegans CDNA  
yk37c3.5; similar to intermediate filament protein"  
/codon\_start=1  
/protein\_id="AA093408.1"  
/db\_xref="pid:g1203951"  
/db\_xref="gi:1203951"  
/translation="TMEASAPQAPMGSGRVLKMTVMGSAIVGIGSPALSAANAKSFL  
EATDKKKTLOGLDRLGNITDRYKLEONRKLIVADIDELRGKWKDTSPIKIVSE  
SLSTARKDIDDAARAKAEVYKVAHRDDLAERYSRVEDIOORRESDEKTSQRTNAT  
ADQSEVEMLRNKRQUTDEKRYTADNSRIWELOKARSDIDETIGRIDQONOVOT  
LMELEFLRVHDEVEVIELQALQAPADTRREFKRNELALAIIDIKEDYDIIAQOGQ  
DMESWYKLVSEVQESANRANMESTYORDEVKRMNDNIGDLRGKLDLENKNSLIEK  
VQNTYQUTDOQRYEALNDPRATLRMRRECOITLVAELQALDITKQMLDAEIIAYR  
KMLEGETRVLQOMVPOAVKTSILOOQENDSTRSVNGEVSTKTTFQSAKGNVTIS  
ECDPENGKFIENSHRNKNDVNGEHI  
DOGINNPPESLYFDGENTWIGANVYTRGKNTLTIYADOGGINNPPESLYFDGENTWIGANV  
complement(28559, .30536)  
/gene="F10c1.5"  
complement(join(28559, .28773,28831, .29052,29467, .29559,  
29871, .30035,30398, .30536))  
/gene="F10c1.5"  
/note="similar to D. melanogaster doublesex protein  
(SP:DSXK\_DROME, P23023)"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AA093409.1"  
/db\_xref="pid:g1203952"  
/db\_xref="gi:1203952"  
/translation="MPVSPSSVSVLPPTSLFQILIRIARNORTPKACRNNHTTGA  
NLGKRRYCOMDKCACCTLIABERQRYAAQVALRROSOBERPARDILEVLGSAGNA  
LDLIDIRLDGEGHGNKHAAPTNNNNTEBEKEDSOGSSPSTGSEVTVSS  
PPELSSESTPNTSGSVITPTSPVFKDNGSSMPKMPNVPYGRPFMRPMPHPFGRPSMS  
DFAASIAAPAFSAQPTIATISSPSTIIPQTAPLDYSHRLMNEAEERYENDXT  
F"  
complement(29306, .29389)  
/gene="F10c1.5"  
/note="Leu: codon recognized: TTA"  
30597, .30788  
/gene="F10c1.1"  
30597, .30788  
/gene="F10c1.1"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AA093410.1"  
/db\_xref="pid:g1203953"  
/db\_xref="gi:1203953"

Query Match 8.5%; Score 19; DB 36; Length 40014;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 GCACCCCAATGTGCCAGAA 159  
DB 37197 GCACCCCAATGTGCCAGAA 37179

RESULT 12  
AC004595/c AC004595 136170 bp DNA PRI 03-FEB-1999  
LOCUS  
DEFINITION Homo sapiens BAC clone RG023M10 from 7p15.3-p21, complete sequence.  
ACCESSION AC004595  
NID 93063514  
VERSION AC004595.1 GI:3063514

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

HTG.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 136170)  
Ozerky, P. and Holmes, A.  
The sequence of Homo sapiens BAC clone RG023M10  
Unpublished (1999)  
2 (bases 1 to 136170)  
Waterston, R.  
Direct Submission  
Submitted (18-APR-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
3 (bases 1 to 136170)  
Waterston, R.  
Direct Submission  
Submitted (03-FEB-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
http://genome.wustl.edu/gsc  
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone; and the assembly was  
confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
http://www.nhgri.nih.gov/DIR/CMB/CHR7, send  
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
Clone RG023M10 is from the first release of the human BAC library  
CITB-978SK-B. The library contains cloned DNA from the male  
fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl.  
Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8  
(1996). This clone is available from Research Genetics, Inc.  
(http://www.resgen.com).  
VECTOR: pBelOBAC11  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RG332B22. Actual start of this  
clone is at base position 1 of RG023M10; actual end is at 136170 of  
RG023M10.

FEATURES  
SOURCE  
1. 136170  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7p15.3-p21"  
/clone\_lib="CITB-978SK-B"  
/clone="RG023M10"  
33. 161  
repeat\_region  
/rpt\_family="MER2L\_g"

repeat\_region 174. .269  
/rpt\_family="Alu"  
repeat\_region 285. .421  
/rpt\_family="MER2L\_g"  
repeat\_region 439. .725  
/rpt\_family="L2"  
repeat\_region 772. .914  
/rpt\_family="L2"  
repeat\_region 1177. .1879  
/rpt\_family="MER50"  
repeat\_region 2951. .3191  
/rpt\_family="MIR"  
repeat\_region 3325. .3373  
/rpt\_family="MIR"  
repeat\_region 3936. .4161  
/rpt\_family="MIR"  
repeat\_region 4869. .5157  
/rpt\_family="Alu"  
repeat\_region 5264. .5396  
/rpt\_family="MER1\_type"  
repeat\_region 5879. .6010  
/rpt\_family="MIR"  
repeat\_region 6115. .6401  
/rpt\_family="Alu"  
repeat\_region 6478. .6638  
/rpt\_family="Achoho"  
repeat\_region 6907. .7067  
/rpt\_family="MER2\_type"  
repeat\_region 7069. .7210  
/rpt\_family="Alu"  
repeat\_region 7211. .7278  
/rpt\_family="MER2\_type"  
repeat\_region 7297. .7637  
/rpt\_family="MER2\_type"  
repeat\_region 8967. .9062  
/rpt\_family="MIR"  
repeat\_region 9609. .9702  
/rpt\_family="L1"  
repeat\_region 10029. .10069  
/rpt\_family="MER4\_group"  
repeat\_region 10728. .10796  
/rpt\_family="L2"  
repeat\_region 10958. .11228  
/rpt\_family="Alu"  
repeat\_region 11349. .11649  
/rpt\_family="Alu"  
repeat\_region 11659. .11855  
/rpt\_family="L1"  
repeat\_region 12576. .12770  
/rpt\_family="MIR"  
repeat\_region 12917. .13234  
/rpt\_family="Alu"  
repeat\_region 14060. .14361  
/rpt\_family="Alu"  
repeat\_region 14656. .15012  
/rpt\_family="Alu"  
repeat\_region 15225. .15709  
/rpt\_family="MER4\_group"  
repeat\_region 15795. .15926  
/rpt\_family="MALR"  
repeat\_region 15985. .16267  
/rpt\_family="Alu"  
repeat\_region 16272. .16331  
/rpt\_family="MALR"  
repeat\_region 16465. .16729  
/rpt\_family="Alu"  
repeat\_region 17696. .18961  
/note="CpG island (AGC-69.2, o/e=0.80, #CpGs=125)"  
18919. 101936  
/gene="Sp4"  
/gene="Sp4"  
CDS  
join(18919. .18925,19344. .19459,19956. .21510,67746. .67974,  
72591. .72790,101689. .101936)  
/gene="Sp4"

```

/note="match to Q02446 (PID:g730781); H_R6023M10.1"
/codon_start=1
/product="SP4 transcription factor"
/protein_id="AAD12226.1"
/db_xref="pid:g4218030"
/db_xref="gi:4218030"
/translation="MSDQKKEEEPEEAAAAMATBEGRKTSPEENNNKRRKTSQSDSO
PSPDLALATCKSIGPGENQATGQQQIITIDPSQGLVQVQNOPOQELVTTDLAGNAM
QIVASTPSPKSNVNSQPASSSSSSSSSSSSSSSSSSSSSSSPGQOVQVONPS
GSVOYVLPOLQTEVQOQIPIPTSSSSSSSSSSSSSSSSSSSSSSSPGQOVQVONPS
ILAOINQVAVQVQIRPVGISPILOTLPTGQAOVYVTPINIGGVTTLALPYINWAA
GGTGVOGQPAFADSGTNGNOIVSTPNTTSSSTMPESPSSSTCTCTTASTLTS
SPTVSSADTGOYASTSASSSRTTETSEPTPATESHAOSSQOLQPNQNOQDOSNS
LQOVIVGQPILOQIOLQOPOQIITQALPQSFQASQGTITITQOQLQVQDAVN
QIAPVAVGAPITLNTAQVSNLOQVONISLSMLQVONAGLSQOLITTPVSSGGTTLA
GVKVOQATAPVAVGIVANATIGAVSPDLOVHLQOQOITSDQVQGRRLRVA
CSCPNCREGEGSNPQPKKHICHIEGCGVYGTSLRAHTRHRTGPERPCTNM
FCGKRRTREDELQRRHRTTGKRCRCPSCSRFMSDLSKHYKTHQKKGCTALA
ITVSELDSVTEVLESPRIVVAISODSNPATPVSTNMEEF"
repeat_region
/rpt_family="MER1_type"
22695..22858
23289..23360
repeat_region
/rpt_family="MER1_type"
23379..23671
24326..24630
repeat_region
/rpt_family="Alu"
25274..25539
26568..26864
repeat_region
/rpt_family="Alu"
27534..27584
repeat_region
/rpt_family="MIR"
29307..29796
repeat_region
/rpt_family="MER2_type"
29985..30167
30276..30574
repeat_region
/rpt_family="Alu"
30810..30985
repeat_region
/rpt_family="MIR"
31235..31535
repeat_region
/rpt_family="Alu"
32706..33004
repeat_region
/rpt_family="Alu"
33223..33349
repeat_region
/rpt_family="L2"
33401..33701

Query Match
Best Local Similarity 8.1%; Score 18; DB 11; Length 136170;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ATTCTTCACCTTCTAGAA 75
Db 38184 ATTCTTCACCTTCTAGAA 38167

RESULT 13
AC004921 150332 bp DNA PRI 14-JAN-1999
DEFINITION Homo sapiens PAC clone DJ0899E09 from Tq11.23-q21.1, complete
sequence.
ACCESSION AC004921
NID 94156176
VERSION AC004921.1 GI:4156176
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 150332)

```

```

AUTHORS Abbott,A., Graves,T. and Stromatt,C.
TITLE The sequence of Homo sapiens PAC clone DJ0899E09
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 150332)
AUTHORS Watson,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 150332)
Waterston,R.
Direct Submission
Submitted (14-JAN-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 14, 1999 this sequence version replaced gi:3213076.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

COMMENT
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by
Pierer de Jong and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu) using the method described by
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
one male donor.
The clone may be obtained either from Genome Systems, Inc.
(http://www.genomesystems.com) or Research Genetics, Inc.
(http://www.resgen.com); or from Pierer de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
This clone sequenced to the right is DJ0562A11. Actual start of
this clone is at base position 1 of DJ0899E09; actual end is at
150332 of DJ0899E09.
Location/Qualifiers
1..150332
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.23-q21.1"
/clone.lib="RPCI-5"
/clone="DJ0899E09"
5..165
/rpt_family="MIR"
295..347
/rpt_family="MIR"
348..647
/rpt_family="Alu"

```

```

repeat_region 651. 671
/rpt_family="(TTAA)n"
repeat_region 672. 780
/rpt_family="MIR"
repeat_region 1001. 1037
/rpt_family="AT_rich"
repeat_region 1327. 1397
/rpt_family="MIR"
repeat_region 1622. 1790
/rpt_family="Alu"
repeat_region 1888. 2193
/rpt_family="Alu"
repeat_region 2288. 2397
/rpt_family="(TA)n"
repeat_region 2468. 2689
/rpt_family="L1"
repeat_region 3106. 3158
/rpt_family="MALR"
repeat_region 3404. 3559
/rpt_family="L1"
repeat_region 3584. 3883
/rpt_family="Alu"
repeat_region 4080. 4308
/rpt_family="L1"
repeat_region 4385. 4410
/rpt_family="AT_rich"
repeat_region 4536. 4582
/rpt_family="(CA)n"
repeat_region 4583. 4703
/rpt_family="Other"
repeat_region 5765. 5808
/rpt_family="(TAA)n"
repeat_region 5810. 6095
/rpt_family="Alu"
repeat_region 7275. 7549
/rpt_family="Alu"
repeat_region 7835. 7863
/rpt_family="(TA)n"
repeat_region 8376. 8431
/rpt_family="(TA)n"
repeat_region 8605. 8850
/rpt_family="L2"
repeat_region 8880. 9243
/rpt_family="MALR"
repeat_region 9244. 10826
/rpt_family="MALR"
repeat_region 10829. 11195
/rpt_family="MALR"
repeat_region 11517. 11539
/rpt_family="AT_rich"
repeat_region 12021. 12309
/rpt_family="Alu"
repeat_region 12791. 12810
/rpt_family="(CAAA)n"
repeat_region 12811. 13053
/rpt_family="Alu"
repeat_region 13072. 13226
/rpt_family="MER2_type"
repeat_region 13227. 13267
/rpt_family="(TA)n"
repeat_region 13268. 13658
/rpt_family="MER2_type"
repeat_region 13659. 13967
/rpt_family="Alu"
repeat_region 13968. 13991
/rpt_family="MER2_type"
repeat_region 13992. 14016
/rpt_family="(TAAA)n"
repeat_region 14017. 16070
/rpt_family="MER2_type"
repeat_region 16073. 16129
/rpt_family="(CA)n"
repeat_region 16130. 16413

```

```

repeat_region /rpt_family="Alu"
16510. 16590
/rpt_family="MIR"
repeat_region 16646. 16946
/rpt_family="Alu"
repeat_region 17613. 17914
/rpt_family="Alu"
repeat_region 18609. 19839
/rpt_family="L1"
repeat_region 19845. 20141
/rpt_family="Alu"
repeat_region 20142. 20442
/rpt_family="Alu"
repeat_region 20927. 21128
/rpt_family="Alu"
repeat_region 21129. 21436
/rpt_family="Alu"
repeat_region 21437. 21609
/rpt_family="Alu"
repeat_region 22656. 22958
/rpt_family="Alu"
repeat_region 23066. 23370
/rpt_family="Alu"
repeat_region 23476. 23778
/rpt_family="Alu"
repeat_region 24000. 24329
/rpt_family="L2"
repeat_region 24858. 25046
/rpt_family="L1"
repeat_region 25064. 25181
/rpt_family="MER86"
repeat_region 25270. 25531
/rpt_family="Alu"
repeat_region 25539. 25819
/rpt_family="Alu"
repeat_region 25895. 25940
/rpt_family="AT_rich"
repeat_region 26111. 26389
/rpt_family="Alu"
repeat_region 26518. 26538
/rpt_family="AT_rich"
repeat_region 26973. 27046

```

```

Query Match 8.1% Score 18; DB 11; Length 150332;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 155 CAGAGATGAAACACTCA 172
|||||
Db 47592 CAGAGATGAAACACTCA 47609

```

```

RESULT 14
AC006210/C 186986 bp DNA PRI 31-DEC-1998
LOCUS Homo sapiens XP22-150 BAC GSHB-309P15 (Genome Systems Human BAC
DEFINITION Library) complete sequence.
ACCESSION AC006210
NID 94090180
VERSION AC006210.1 GI:4090180
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 186986)
Muzny,D., Arenson,A.D., Bouck,J., Brundage,E., Bunac,C., Chen,Z.,
Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C.,
Gorell,J.H., Gorrell,L.L., Hernandez,J., Jackson,L.,
Kondejewski,N., Leal,B., Licharge,O., Liu,W., Logan,O., Lu,J.,
Martinez,C., Oswald,G., Pampell,L.R., Parish,B.J., Perez,L.,
Raschid,N.D., Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,O.,
Williamson,A., Worley,K.C., Yu,W., Zhou,X., Nelson,D. and

```





### Best Local

Mon Aug 9 13:26:41 1999

us-09-049-696-1.rge

Page 17

Matches	18;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	108	CAACAATGGCTATGAAG	125						
Db	7966	CAACAATGGCTATGAAG	7983						

Search completed: August 6, 1999, 11:32:51  
Job time: 6057 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 15:56:05 ; Search time 650.92 Seconds

(without alignments)  
85.714 Million cell updates/sec

Title: US-09-049-696-1

Perfect score: 223  
Sequence: 1 GGAATCAGCAGGAGATGTC.....ATCTCTGATCTGTTTGAG 223

Scoring table: OLIGO\_MUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17	7.6	2287	1 Q13802	Non-receptor like
C 2	17	7.6	2155	1 Q42418	IDUA - exons I and
C 3	17	7.6	4475	1 Q45895	IDUA - exons III t
C 4	17	7.6	13104	1 Q46852	Clone of recombin
C 5	17	7.6	267	1 T20069	Human gene signatu
C 6	17	7.6	2067	1 T71754	Human alpha-L-idur
C 7	17	7.6	1883	1 V52044	Helicobacter poly
C 8	17	7.2	17327	1 Q44278	Serlycln - proteo
C 9	16	7.2	1465	1 Q50871	BSP 14-3-3. New pl
C 10	16	7.2	3020	1 T03275	Mutant human cilia
C 11	16	7.2	3019	1 T03276	Wild type human c
C 12	16	7.2	9534	1 T62072	Vector RP3224E2 en
C 13	16	7.2	1507	1 V04698	Homo sapiens 20q13
C 14	16	7.2	10382	1 V09023	Homo sapiens 20q13
C 15	16	7.2	110000	1 V21209_01	Continuation (2 of
C 16	15	6.7	988	1 N91023	Transforming growt
C 17	15	6.7	972	1 N91881	Nucleotide sequenc
C 18	15	6.7	787	1 N91284	Sequence of insert
C 19	15	6.7	972	1 N82260	Plasmid pRTB4 cont
C 20	15	6.7	787	1 N70522	Sequence of Richin
C 21	15	6.7	806	1 N91891	Sequence of castor
C 22	15	6.7	1446	1 N70526	Sequence of cDNA 1
C 23	15	6.7	1801	1 N70524	Sequence of cDNA 1
C 24	15	6.7	1372	1 N60166	Sequence encoding
C 25	15	6.7	3401	1 N60518	Open reading frame
C 26	15	6.7	897	1 N50160	Sequence encoding
C 27	15	6.7	990	1 N50284	Sequence encoding
C 28	15	6.7	795	1 N50378	Plasmid pRTB-4 ins
C 29	15	6.7	3211	1 N50347	Positive clone (pha
C 30	15	6.7	6063	1 Q37205	Delta-amino leuval
C 31	15	6.7	883	1 Q78793	TCR alpha-chain cl
C 32	15	6.7	9272	1 Q79353	Human genomic clon
C 33	15	6.7	837	1 Q74396	Isoform 1g11 of th
C 34	15	6.7	867	1 Q72873	Human transforming
C 35	15	6.7	1789	1 Q92296	Erythropoietin cDN
C 36	15	6.7	1092	1 Q91360	T-cell receptor al
C 37	15	6.7	696	1 Q91361	T-cell receptor al
C 38	15	6.7	807	1 Q91362	T-cell receptor al
C 39	15	6.7	1430	1 Q92525	Nicotiana glumagl
C 40	15	6.7	10614	1 Q89555	Hamster cholestero
C 41	15	6.7	5394	1 T00872	Murine MC26 struct
C 42	15	6.7	2481	1 T13002	Human adrenergic G
C 43	15	6.7	2894	1 T10423	H. Influenzae SB33

44 15 6.7 1912 1 T17715  
45 15 6.7 1536 1 X13019

## ALIGNMENTS

RESULT 1	
ID Q13802/c	standard; cDNA: 2287 BP.
AC Q13802;	
DT 09-DEC-1991 (first entry)	
DE Non-receptor linked protein tyrosine phosphatase.	
KW PRPase; malignancy; cancer; ss.	
OS Homo sapiens.	
FM Key	Location/Qualifiers
FT cds	61..1308
FT	/tag= a
FT	/product= PRPase
FT	1512..1526
FT	polya_signal
FT	1677..1682
FT	/tag= b
FT	polya_signal
FT	1677..1682
FT	/tag= c
PN MO913989-A.	
PD 19-SEP-1991.	
PF 14-MAR-1991: D01748.	
PR 14-MAR-1990: US-494036.	
PA (WASH-) WASHINGTON RES FOUN.	
PI Fischer EH, Krebs EG, Tonks NK, COOL DE;	
DR WPI: 91-295643/40.	
DR P-PSDB: R14114.	
PT New DNA encoding non-receptor-linked protein tyrosine phosphatase	
PT - used in treating malignancies associated with protein-linked	
PT tyrosine phosphatase.	
PS Clam 1; Fig 1; 38pp; English.	
CC A cDNA library prep. from human peripheral T cell poly(A)+ mRNA was	
CC screened using probes designed from the sequence of human placenta	
CC PRPase 1B. The sequence can be used to express recombinant PRPase	
CC or derivs. useful in cancer therapy where a protein kinase is	
CC involved.	
CC See also Q13817-Q13820.	
SQ Sequence 2287 BP; 753 A; 411 C; 464 G; 659 T;	
Query Match	7.6%; Score 17; DB 1; Length 2287;
Best local Similarity	100.0%; Pred. No. 6;
Matches 17; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 203 CATCTGTATCTGTTT 219	
DB 211 CATCTGTATCTGTTT 195	
RESULT 2	
ID Q42418/c	standard; cDNA: 2155 BP.
AC Q42418;	
DT 27-SEP-1993 (first entry)	
DE IDUA - exons I and II.	
KW Alpha-L-iduronidase; glycosaminoglycan alpha-L-iduronohydrolase;	
OS Homo sapiens.	
FM Key	Location/Qualifiers
FT cds	89..2050
FT	/tag= a
FT	signal_peptide
FT	89..167
FT	/tag= b
FT	mat_peptide
FT	168..2047
FT	/tag= c
FT	exon
FT	90..246
FT	/tag= d
FT	number= 1
FT	247..387
FT	exon



SQ Sequence 4475 BP; 619 A; 1566 C; 1514 G; 776 T;

Query Match 7.6%; Score 17; DB 1; Length 4475;

Best Local Similarity 100.0%; Pred. No. 6.2;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 121 GAAGCATGTCTGTC 137

DB 2167 GAAGCATGTCTGTC 2151

RESULT 4

ID 046852 standard; DNA; 13104 BP.

AC 046852;

DT 26-JAN-1994 (first entry)

DE Clone of recombinant human kappa casein gene fragment.

KM Casein; supplement; milk; pharmaceutical; ss.

OS Homo sapiens. Location/Qualifiers

FT Key 1.8834

FT Intron

FT Exon

FT Exon

FT Intron

FT Exon

FT Exon

FT Intron

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

FT Exon

PN MO9514772-A1.

PD 11-NOV-1995.

PE 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PI (OKUB/) OKUBO K.

PT Matsubara K. Okubo K.

PS WPI; 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS claim 1: Page 552; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in 119001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

CC Sequence 267 BP; 82 A; 33 C; 47 G; 94 T;

Query Match 7.6%; Score 17; DB 1; Length 267;

Best Local Similarity 100.0%; Pred. No. 5.6;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 ATCTTCACCTCTAGA 74

DB 19 ATTCTTCACCTCTAGA 3

RESULT 6

ID T71754/c

AC T71754; standard; cDNA; 2067 BP.

DT 01-OCT-1997 (first entry)

DE Human alpha-L-iduronidase cDNA.

KM Lysosomal enzyme; lysosome; transgenic plant; alpha-L-iduronidase;

OS Homo sapiens. Location/Qualifiers

FT Key 1.2067

FT cds

FT /tag= a

FT /transl\_except= (pos:97..99, aa:His)

FT /transl\_except= (pos:841..843, aa:Gln)

PN MO9710353-A1.

PD 20-MAR-1997.

PE 13-SEP-1996; U14730.

PR 14-SEP-1995; US-003737.

PA (CROP-) CROPTTECH DEV CORP.

PI (VIRG ) VIRGINIA TECH INTELLECTUAL PTY INC.

PT Cramer CL, Oishi KK, Radin DN, Weissensorn DL.

PS WPI; 97-202248/18.

PT Production of enzymatically active (modified) lysosomal enzyme in

PT transgenic plants - useful in treatment of lysosomal storage

PT disorders

PS Example 2: Page 65-66; 11pp; English.

CC A cDNA sequence (T71754) codes for human alpha-L-iduronidase

CC (M1838). It was obtd. by PCR amplification using primers (see

CC also T71757-58) designed to incorporate restriction sites to

CC facilitate Mega cloning. The construct, placed under control of the

CC inducible Mega promoter (see also T71752), can drive expression of

CC hCG in transgenic (esp. tobacco) plants. The plant expression

CC system provides for post-translational modification and processing  
 CC to produce enzymatically active iduronidase, which can be harvested  
 CC for use in Hurler syndrome enzyme replacement therapy.  
 SO Sequence 2067 BP; 318 A; 776 C; 635 G; 338 T;

Query Match 7.6%; Score 17; DB 1; Length 2067;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 121 GAAGCATTGCTGTC 137  
 ||||||||||||||||  
 DB 1056 GAAGCATTGCTGTC 1040

RESULT 7  
 VS2044

ID VS2044 standard; DNA; 1883 BP.

AC VS2044.

DT 09-NOV-1998 (first entry)

DE Helicobacter polypeptide GHPO 1284 DNA.

KM GHPO 1284; infection; therapy; diagnosis; vaccine; gastritis;

KW ulcer; ss.

OS Helicobacter pylori.

FT Key Location/Qualifiers

FT CDS 91..1836

FT /tag= a

FT MO981225-A1.

PD 22-MAY-1998.

PR 14-NOV-1997; U21353.

PR 29-JUL-1997; US-902615.

PR 14-NOV-1996; US-749051.

PR 01-APR-1997; US-831309.

PR 01-APR-1997; US-833457.

PR 24-JUN-1997; US-881227.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (PLNC) MAX PLANCK GES FORSCHUNG WISSENSCHAFTEN.

PA (INNR) MERIEUX ORANVAX PASTEUR MERIEUX SERUMS.

PI A1-garawi A, Haas R, Kleanthous H, Meyer T, Miller C,

PI Odenbreit S, Tomb J;

DR WPI: 98-297855/26.

DR P-PEDB: W71509.

PT Helicobacter polynucleotide and polypeptide sequences - useful to

PT treat or prevent gastrointestinal infection

PS Disclosure; Page 176-179; 362pp; English.

CC This polynucleotide encodes Helicobacter pylori polypeptide

CC GHPO 1284 (see W71509). 85 Polynucleotides (see V52009-V52093)

CC encoding claimed Helicobacter polypeptides (see W71474-W71558) have

CC been identified in the H. pylori genome using a transposon shuttle

CC mutagenesis method in which Tmmax9 mini-blam was used for

CC insertional mutagenesis of an H. pylori gene library established in

CC Escherichia coli. The invention provides: methods for producing

CC these polypeptides in recombinant host systems, and related

CC expression cassettes, vectors and transformed or transduced host

CC cells; live vaccine vectors that contain polynucleotides of the

CC invention and which can be used to prevent or treat Helicobacter

CC infection; therapeutic and/or prophylactic methods involving

CC administration of polynucleotide molecules (either in naked form

CC or delivered with a delivery vehicle), polypeptides or monospecific

CC antibodies; methods for detecting the presence of Helicobacter

CC in samples using e.g. polynucleotide molecules; and methods for

CC purifying polypeptides of the invention.

SO Sequence 1883 BP; 659 A; 296 C; 415 G; 512 T;

Query Match 7.6%; Score 17; DB 1; Length 1883;

Best Local Similarity 100.0%; Pred. No. 6;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 149 ATGTGCCAGAGATGAA 155  
 ||||||||||||||||  
 DB 953 ATGTGCCAGAGATGAA 969

RESULT 8

O44278/c

ID O44278 standard; DNA; 17327 BP.

AC O44278;

DT 24-NOV-1993 (first entry)

DE Serglycin - proteoglycan peptide core.

KM Haematopoietic cell secretory granule proteoglycan;

KW positive; negative; transcriptional regulatory element;

KW enhancer; eukaryotic promoter; constitutive suppressor;

KW TATA-box; ss.

OS Homo sapiens.

FT Key

FT misc\_signal

FT Location/Qualifiers

FT 373..621

FT /tag= a

FT /note= "5' regulatory region;

FT claim 1-3, page 78"

FT 373..433

FT /tag= b

FT /note= "negative transcriptional regulatory element;

FT claim 6 and 9-10, page 78-79"

FT 514..541

FT /tag= c

FT /note= "enhancer transcriptional regulatory element;

FT claim 11 and 14-15, page 79"

FT 582..602

FT /tag= d

FT /note= "eukaryotic promoter element;

FT claim 16 and 19-20, page 80"

FT 675..16646

FT /tag= e

FT /label= serglycin

FT 622..753

FT /tag= f

FT /number= 1

FT 9597..9744

FT /tag= g

FT /number= 2

FT 16397..17327

FT /tag= h

FT /number= 1

FT 675..9598

FT /tag= i

FT /note= "interrupted by exon 1"

FT 16449..16502

FT /tag= j

FT /note= "serine-glycine rich glycosaminoglycan

FT attachment region

FT 17062..17067

FT /tag= k

FT misc\_difference

FT 6135

FT /tag= l

FT /note= "base represented as N in the specification"

PN WO9313119-A.

PD 08-JUL-1993.

PF 23-DEC-1992; U11194.

PR 03-JAN-1992; US-816289.

PR 02-JUL-1992; US-906871.

PA (BGM) BRIGHAM & WOMEN'S HOSPITAL.

PI Avraham S, Stevens RL;

DR WPI: 93-227261/28.

DR P-PEDB: R39393.

PT Transcription regulatory elements of ser-glycin gene - specific

PT for haematopoietic cells, also trans-acting transcriptional

PT binding factors Fig 4B-J; 112pp; English.

PS Disclosure; Fig 4B-J; 112pp; English.

CC A negative transcription regulatory element (a constitutive

CC suppressor) between residues -250 and -190 (see tag b) of the 5'

CC flanking region of the human serglycin gene, a positive

CC (hematopoietic cell enhancer) regulatory element located between

CC residues -118 and -81 (see tag c), an equiv. of the TATA-box and

CC a novel eukaryotic promoter that utilizes such equiv. (see tag d)



CC are identified. The regulatory elements, vectors and hosts  
 CC provided with these elements, are useful in the control of gene  
 CC transcription of heterologous genes in eukaryotic cells, esp.  
 CC hematopoietic cells.  
 SQ Sequence 17327 BP; 4936 A; 3604 C; 3741 G; 5045 T;

Query Match 7.2%; Score 16; DB 1; Length 17327;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 CACTATTCAGCTGAA 107  
 |||||  
 DB 15259 CACTATTCAGCTGAA 15244

RESULT 9  
 ID 050871/c  
 AC 050871 standard; CDNA; 1465 BP.  
 DT 13-MAY-1994 (first entry)  
 DE BSP 14-3-3.  
 KW Brain specific protein; BSP; plant cell; callus; activator;  
 KW tyrosine monooxygenase; tryptophan monooxygenase; vector; ss.  
 OS Oryza sativa.  
 FH Key  
 FT cds Location/Qualifiers  
 FT 243..1022  
 FT /tag= a  
 FT /product= BSP14-3-3\_protein\_equivalent  
 PN J05260975-A.  
 PD 12-OCT-1993.  
 PF 23-MAR-1992; 065299.  
 PR (POK) POLA CHEM IND INC.  
 PA WPI; 93-356451/45.  
 DR P-PSDB; R42198.  
 PT New plant gene with specified aminoacid sequence - codes protein  
 PT having functions equiv. to those of brain specific portion  
 PS Disclosure; Page 6-7; 12pp; Japanese.  
 CC The sequence encodes a plant protein having functions specific to  
 CC that of brain specific protein BSP14-3-3. The expression of this  
 CC gene in cultured plants cells, callus etc. brings forth activators  
 CC of enzymes e.g. tyrosine monooxygenase and tryptophan  
 CC monooxygenase.  
 CC Sequence 1465 BP; 366 A; 332 C; 387 G; 380 T;  
 SQ

Query Match 7.2%; Score 16; DB 1; Length 1465;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 177 ACAATTAAGACATG 192  
 |||||  
 DB 1340 ACAATTAAGACATG 1325

RESULT 10  
 ID T03275/c  
 AC T03275 standard; DNA; 3020 BP.  
 DT 09-APR-1996 (first entry)  
 DE Mutant human ciliary neurotrophic factor gene.  
 KW Wild type; human; ciliary neurotrophic factor; CNF; Intron; frameshift;  
 KW splice acceptor site; open reading frame; keyhole limpet haemocyanin;  
 KW antibody; nerve disease; amyotrophic lateral sclerosis; mutant;  
 KW senile dementia; progressive motor neuropathy;  
 KW peripheral nerve interference; ds.  
 OS Synthetic.  
 FH Key  
 FT cds Location/Qualifiers  
 FT 46..1398  
 FT /tag= a  
 FT /product= mutant human ciliary neurotrophic factor  
 FT 46..162  
 FT exon

FT /tag= b  
 FT /note= "exon 1"  
 FT 163..1323  
 FT /tag= c  
 FT exon 1324..1395  
 FT /tag= d  
 FT /note= "exon 2"

PN J07203973-A.  
 PD 08-AUG-1995.  
 PF 11-JAN-1994; 001265.  
 PR 11-JAN-1994; JP-001265.  
 PA (BMLB-) BML KK.  
 PA (TOKS-) TOKYO SHINKI KAGAKU SOGO KENYUSHO ZH.  
 DR WPI; 95-307168/40.  
 DR P-PSDB; R862989.  
 PT Human ciliary neurotrophic factor mutant sequence - useful for  
 PT establishing genotype associated with diseases of the nervous system  
 PT including senile dementia and amyotrophic lateral sclerosis  
 PS Example 1; Page 5-6; 9pp; Japanese.  
 CC The nucleotide sequence of a gene encoding a mutant human ciliary  
 CC neurotrophic factor (CNF). The gene contains one Intron. The mutant  
 CC gene was constructed by altering the sequence at the 3' end of the Intron  
 CC in the wild type gene (T03276) from GGCC to AGCC. This mutation creates  
 CC an alternative 3' splice acceptor site which frameshifts the open reading  
 CC frame by 1 base and causes the reading frame to terminate prematurely  
 CC after a further 24 amino acids. The new shorter coding sequence (T03274)  
 CC encodes a protein of 62 amino acids (R86298) instead of 200 amino acids.  
 CC The C-terminal 22 amino acids of the mutant protein sequence were linked  
 CC to keyhole limpet haemocyanin used to raise antibodies to the mutant  
 CC protein. These antibodies and the mutant gene can be used to detect  
 CC various nerve diseases such as senile dementia, amyotrophic lateral  
 CC sclerosis, progressive motor neuropathy and peripheral nerve  
 CC interference.  
 CC Sequence 3020 BP; 843 A; 600 C; 660 G; 917 T;  
 SQ

Query Match 7.2%; Score 16; DB 1; Length 3020;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 109 AACATGCTATGAG 124  
 |||||  
 DB 505 AACATGCTATGAG 490

RESULT 11  
 ID T03276/c  
 AC T03276 standard; DNA; 3019 BP.  
 DT 09-APR-1996 (first entry)  
 DE Wild type human ciliary neurotrophic factor gene.  
 KW Wild type; human; ciliary neurotrophic factor; CNF; Intron; frameshift;  
 KW splice acceptor site; open reading frame; keyhole limpet haemocyanin;  
 KW antibody; nerve disease; amyotrophic lateral sclerosis; mutant;  
 KW senile dementia; progressive motor neuropathy;  
 KW peripheral nerve interference; ds.  
 OS Homo sapiens.  
 FH Key  
 FT cds Location/Qualifiers  
 FT 46..1815  
 FT /tag= a  
 FT /product= human ciliary neurotrophic factor  
 FT 1408..1410  
 FT /tag= b  
 FT /transl\_except= Seq: GAG, a.a.:TTP  
 FT exon 46..162  
 FT /tag= c  
 FT /note= "exon 1"  
 FT Intron 163..1326  
 FT /tag= d  
 FT exon 1327..1812  
 FT /tag= e  
 FT /note= "exon 2"  
 FT PN J07203973-A.

PD 08-AUG-1995.  
PF 11-JAN-1994; 001265.  
PR 11-JAN-1994; JP-001265.  
PA (BMLB-) BML KK.  
PA (TOKS-) TOKYO SHINKAI KAGAKU SOGO KENKYUSHO ZH.  
DR WPI; 95-307168/40.  
DR P-PSDB; R86298.  
PT Human ciliary neurotrophic factor mutant sequence - useful for  
PT establishing genotype associated with diseases of the nervous system  
PS including senile dementia and amyotrophic lateral sclerosis  
PS Disclosure: Page 6-8; 9pp; Japanese.  
CC The nucleotide sequence of the wild type gene encoding the human ciliary  
CC neurotrophic factor (CNTF). The gene contains one intron. The gene was  
CC mutated by altering the sequence at the 3' end of the intron from GGCC to  
CC AGCC. This alteration creates an alternative 3' splice acceptor site  
CC which frameshifts the protein reading frame by 1 base and causes the  
CC reading frame to terminate prematurely after a further 24 amino acids  
CC (see T03275). The new shorter coding sequence (T03274) encodes a protein  
CC of 62 amino acids (R86298) instead of 200 amino acids. The C-terminal 22  
CC amino acids of the mutant protein sequence were linked to keyhole limpet  
CC haemocyanin used to raise antibodies to the mutant protein. These  
CC antibodies and the mutant gene can be used to detect various nerve  
CC diseases such as senile dementia, amyotrophic lateral sclerosis,  
CC progressive motor neuropathy and peripheral nerve interference.  
CC Sequence 3019 BP; 842 A; 598 C; 661 G; 918 T;  
SQ

Query Match 7.2%; Score 16; DB 1; Length 3019;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 AACATGGCTATGAAG 124  
DB 505 AACATGGCTATGAAG 490  
|||||

RESULT 12  
T62072/c  
ID T62072 standard; DNA; 9534 BP.  
AC T62072.  
DT 06-JUL-1997 (first entry)  
DE Vector RP3224E2 encoding human ciliary neurotrophic factor.  
KW CNTF; ciliary neurotrophic factor; central nervous system;  
KW growth factor; amyotrophic lateral sclerosis; therapy; vector;  
KW plasmid RP3224E2; BHK; encapsulation; ss.  
OS Chimeric Homo sapiens;  
OS Chimeric Mus sp.;  
OS Chimeric herpes simplex virus.  
OS WO9712635-A1.  
PN 10-APR-1997.  
PF 02-OCT-1996; U15824.  
PR 02-OCT-1995; US-537338.  
PA (CYTO-) CYTOTHERAPEUTICS INC.  
PA Hebscher P, Baetge EE, Hamman JP;  
PI WPI; 97-225980/20.  
DR Direct administration of ciliary neurotrophic factor into central  
PT nervous system - to treat amyotrophic lateral sclerosis  
PS Example: Page 52-55; 74pp; English.  
CC Vector RP3224E2 (T62072) is a pMTT-based vector contg. the human  
CC ciliary neurotrophic factor (hCNTF) gene fused to a murine  
CC immunoglobulin signal peptide (see also W13388) under control of  
CC the murine metallothionein promoter, and also contg. the herpes  
CC simplex virus thymidine kinase gene at the NotI site. The vector  
CC was used to transfect baby hamster kidney (BHK) cells, allowing  
CC prodn. of biologically active hCNTF. These BHK cells can be  
CC encapsulated and administered to the central nervous system for  
CC treatment of amyotrophic lateral sclerosis.  
SQ Sequence 9534 BP; 2260 A; 2478 C; 2513 G; 2283 T;

Query Match 7.2%; Score 16; DB 1; Length 9534;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 AACATGGCTATGAAG 124  
DB 1220 AACATGGCTATGAAG 1205  
|||||

RESULT 13  
V04698  
ID V04698 standard; CDNA; 1507 BP.  
AC V04698;  
DT 21-JUL-1998 (first entry)  
DE Homo sapiens 20q13 amplicon cc49 transcript.  
KW 20q13 amplicon; chromosome 20; tumour; detection; C2H2 zinc finger;  
KW chromosomal abnormalities; probe; gene therapy; antisense inhibition;  
KW treatment; age-related macular degeneration; retinitis pigmentation;  
KW Leber's congenital amaurosis; ds.  
OS Homo sapiens.  
PN WO9802539-A1.  
PD 22-JAN-1998.  
PF 15-JUL-1997; U12343.  
PR 17-JAN-1997; US-785332.  
PR 15-JUL-1996; US-680395.  
PR 16-OCT-1996; US-731499.  
PA (REGC) UNIV CALIFORNIA.  
PI Collins CC, Godfrey T, Gray JW, Hwang SI, Kowbel D,  
PI Rommens J;  
DR WPI; 98-110587/10.  
PT New sequences from the 20q13 amplicon - used for detecting  
PT chromosomal abnormalities, particularly tumours, and for developing  
PT products for treating diseases  
PS Claim 1: Pages 61-62; 91pp; English.  
CC The sequence is that of a cDNA sequence cc49, which was isolated  
CC from the 20q13 amplicon. It shows homology to C2H2 zinc finger  
CC genes. It can be used as a probe for the detection  
CC of chromosomal abnormalities at 20q13. It and other  
CC sequences isolated from the 20q13 amplicon are consistently  
CC amplified in primary tumours. These sequences are useful as probes  
CC or as probe targets for monitoring the relative copy number of  
CC corresponding sequences from a biological sample such as tumour  
CC cells. The sequences can also be used in therapeutic applications  
CC for modulating the expression of the endogenous gene or the activity  
CC of the gene product. Examples of therapeutic approaches include  
CC antisense inhibition of gene expression, gene therapy, and monoclonal  
CC antibodies that specifically bind the gene products. The products can  
CC also be used in the treatment of other diseases, e.g. age-related  
CC macular degeneration, Leber's congenital amaurosis and retinitis  
CC pigmentation.  
CC Sequence 1507 BP; 445 A; 354 C; 364 G; 340 T;  
SQ

Query Match 7.2%; Score 16; DB 1; Length 1507;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 TGAACACTCATTCOA 177  
DB 32 TGAACACTCATTCOA 47  
|||||

RESULT 14  
V09023  
ID V09023 standard; DNA; 10282 BP.  
AC V09023;  
DT 21-JUL-1998 (first entry)  
DE Homo sapiens 20q13 amplicon ZABC-1 genomic sequence.  
KW 20q13 amplicon; chromosome 20; tumour; detection; ZABC-1 gene;  
KW chromosomal abnormalities; probe; gene therapy; antisense inhibition;  
KW treatment; age-related macular degeneration; retinitis pigmentation;  
KW Leber's congenital amaurosis; zinc finger amplified in breast cancer;  
KW ds.  
OS Homo sapiens.  
PN WO9802539-A1.  
PD 22-JAN-1998.

DB 45316 TCGTATCTGTTGAA 45331

Search completed: August 5, 1999, 17:39:19  
Job time: 6194 sec

PF 15-JUL-1997: U12343.  
PR 17-JAN-1997: US-785532.  
PR 15-JUL-1996: US-680395.  
PR 16-OCT-1996: US-731499.  
PA (REGC ) UNIV CALIFORNIA.  
PI Collins CC, Godfrey T, Gray JW, Hwang SI, Kowbel D,  
PI Rommens J:  
DR WPI: 98-110587/10.  
DR P-PSDB: W23975.  
PI New sequences from the 20q13 amplicon - used for detecting  
PI chromosomal abnormalities, particularly tumours, and for developing  
PI products for treating diseases  
PS Claim 1: Pages 64-67; 91pp: English.  
CC The sequence is that of the genomic sequence of ZABC-1 (zinc finger  
CC amplified in breast cancer). It maps to the core of the 20q13.2  
CC amplicon and is overexpressed in primary tumours and breast cancer  
CC cell lines having 20q13.2 amplification. The exact coding region for  
CC the genomic sequence is not given. The sequence can be used as a  
CC probe for the detection of chromosomal abnormalities at 20q13. It and  
CC other sequences isolated from the 20q13 amplicon are consistently  
CC amplified in primary tumours. These sequences are useful as probes  
CC or as probe targets for monitoring the relative copy number of  
CC corresponding sequences from a biological sample such as tumour  
CC cells. The sequences can also be used in therapeutic applications  
CC for modulating the expression of the endogenous gene or the activity  
CC of the gene product. Examples of therapeutic approaches include  
CC antisense inhibition of gene expression, gene therapy, and monoclonal  
CC antibodies that specifically bind the gene products. The products can  
CC also be used in the treatment of other diseases, e.g. age-related  
CC macular degeneration, Leber's congenital amaurosis and retinitis  
CC pigmentatio.

SQ Sequence 10282 BP; 2820 A; 2222 C; 2191 G; 3047 T;

## Query Match

Best Local Similarity 100.0%; Score 16; DB 1; Length 10282;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 TGAACACTCATTCAA 177

DB 2453 TGAACACTCATTCAA 2468

## RESULT 15

V21209\_01 Continuation (2 of 17) of V21209 from base 100001 (Methanococcus jannaschii circular chr

WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

## Query Match

Best Local Similarity 100.0%; Score 16; DB 1; Length 110000;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 TCGTATCTGTTGAA 222

|||||

---

This Page Blank (uspto)

Page Blank (uspto)

Gencore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 04:01:27 ; Search time 921.96 Seconds  
(without alignments)  
74.084 Million cell updates/sec

Title: US-09-049-696-2

Perfect score: 273  
Sequence: 1 GTTGCATTCGACCCCAATGT.....AGATGGCACTGTGGCGAG 273

Scoring table: IDENTITY\_MUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.8	20.1	398	1 X40511	Human secreted pro
2	34.4	12.6	1200	1 Q52267	Sequence encoding
3	33.2	12.2	799	1 X13748	Enterococcus faeca
4	31.8	11.6	1947	1 X08971	Human herpesvirus
5	31.6	11.6	1947	1 X08971	Human herpesvirus
6	31.6	11.6	2980	1 N91071	Nucleotide sequenc
7	31.6	11.6	2615	1 N81002	Bacillus thuringie
8	31.6	11.6	2980	1 N82139	Cryc gene. New gen
9	31.6	11.6	2983	1 Q71026	CryIIIA gene which
10	31.6	11.6	2615	1 X3187	Bacillus thuringie
11	31.2	11.4	110000	1 V21209_01	Confination (2 of
12	30.4	11.1	3475	1 X11720	Food-grade plasmid
13	30.4	11.1	3475	1 V52150	Streptococcus pneu
14	29.8	10.9	9064	1 V21209_15	Confination (16 o
15	29.6	10.8	110000	1 X25885	C. albicans alpha-I
16	29.6	10.8	5194	1 X25885	B. vulgatis H81-Pr
17	29.4	10.7	560	1 X00668	Human secreted pro
18	29.2	10.7	560	1 X20248_01	Confination (2 of
19	29.2	10.7	110000	1 Q60002	Human brain expres
20	29	10.6	372	1 Q75139	C. jejuni hippuric
21	29	10.6	1338	1 Q75139	Fleeced Morning GI
22	29	10.6	6412	1 Q79622	Enterococcus faeca
23	29	10.6	12022	1 X13054	Enterococcus faeca
24	28.8	10.5	32768	1 X13060	Enterococcus faeca
25	28.8	10.5	1661	1 X13564	Human secreted pro
26	28.8	10.5	2361	1 X04316	Human secreted pro
27	28.6	10.5	110000	1 V21209_09	Confination (10 o
28	28.6	10.5	581	1 X21118	Polynucleotide seq
29	28.4	10.4	1230	1 Q45290	Sequence of the re
30	28.4	10.4	3467	1 V21462	Murine protein tar
31	28.4	10.4	1230	1 V48354	Avian infectious b
32	28.4	10.4	2869	1 V52263	Streptococcus pneu
33	28.4	10.4	110000	1 X20248_00	Borrelia burgdorfe
34	28.4	10.4	3393	1 X24681	Human synaptonemal
35	28.2	10.3	110000	1 V21209_03	Confination (4 of
36	28.2	10.3	110000	1 V21209_04	Confination (5 of
37	28.2	10.3	11515	1 X20536	Polynucleotide seq
38	28	10.3	3324	1 T97393	Aspergillus oryzae
39	28	10.3	9072	1 V74336	Staphylococcus aur
40	27.8	10.3	1560	1 X00655	Human secreted pro
41	27.8	10.2	2280	1 T01403	Canine herpesvirus
42	27.8	10.2	1761	1 T01407	H6-promoted CHV GC
43	27.8	10.2	110000	1 T58840_2	Confination (3 of

44 27.8 10.2 6216 1 V06946  
c 45 27.8 10.2 51952 1 V26084

## ALIGNMENTS

RESULT 1  
ID X40511 standard; CDNA: 398 BP.  
AC X40511;  
DE 18-JUN-1999 (first entry)  
KW Human secreted protein 5' EST SEQ ID No: 111.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN W09906350-82.  
PD 11-FEB-1999.  
PE 31-JUL-1998; IB1232.  
PR 01-AUG-1997; US-905144.  
PA (GENSET) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B:  
DR WPI: 99-153780/13.  
DR P-PSDB; Y11789.  
PT New isolated prostate-derived nucleic acids - used to develop  
PT products which may have cytokine, immune regulatory, haematopoiesis  
PT regulating, anti-inflammatory or tumour inhibition activity  
PS Clam J; Page 228; 675pp; English.  
CC X4038 to X40715 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins expressed in prostate, and encode the proteins given in  
CC Y11716 to Y11993 respectively. The proteins given represent the signal  
CC peptide and an N-terminal fragment of a secreted protein. The nucleic  
CC acid sequences can be used for producing secreted human gene products.  
CC They can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell proliferation and  
CC differentiation activity, haematopoiesis regulating activity, tissue  
CC growth regulation activity, reproductive hormone regulating activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition  
CC activity or other activities. The products can be used in forensic, gene  
CC therapy and chromosome mapping procedures. The sequences can also be used  
CC for obtaining corresponding promoter sequences. The nucleic acids  
CC encoding the signal peptides can be used for directing extracellular  
CC secretion of a polypeptide or the insertion of a polypeptide into a  
CC membrane, or importing a polypeptide into a cell.  
SQ Sequence 398 BP; 123 A; 86 C; 76 G; 113 T;

Query Match 20.1%; Score 54.8; DB 1; Length 398;  
Best local similarity 62.3%; Pred. No. 5.7e+08;  
Matches 86; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 2 TTGCATTCGACCCCAATGTGCCAGAGATGAAGAACTATTAAACAATAAGAGCATGG 61  
DB 256 TTGCATTCATTAATCTCAGTACCTGAGATACAGAACTCTCAACATTAAGGAATGA 315  
QY 62 TACCCAGGACCTCTGATATGTTGAGCTACAGGAAGGATTTTTCACAAATG 121  
DB 316 TAATCGAACCTTATTTACCTATTATTAATGTCACAGAGAAAGATATTTTCAGAAATA 375  
QY 122 TTGCATTTGATTCCTG 139  
DB 376 TAAAGATTTTAATCTG 393  
RESULT 2  
ID Q52267  
AC Q52267;  
Q52267 standard; DNA: 1200 BP.

	DE	02-JUN-1994 (first entry)	
	DE	Sequence encoding 5S rRNA gene of Candida albicans.	
	KW	Detection: fungal infection; ribosomal RNA; rRNA; ribonucleic acid;	
	KM	repeat unit; Candida albicans; identification; ds.	
	OS	Candida albicans.	
FH	Key	Location/Qualifiers	
FT	cds	937..1057	
FT	/tag= a		
FT	/product= 5S rRNA		
PN	MO9323568-A.		
PD	25-NOV-1993.		
PR	18-MAY-1993; NZ0039.		
PA	(CANN/) CANNON R D.		
PA	(JENK/) JENKINS A R.		
PA	(SHEP/) SHEPHERD M G.		
PI	Canon RD, Holmes AR, Jenkinson HF, Shepherd MG;		
PT	WPI: 93-386598/48.		
PT	Diagnosis of fungal infections by detecting specific genomic		
PT	regions in pathogenic strains - using probes specific for the 5S		
PT	rRNA gene and/or non transcribed spacer region of the rRNA repeat		
PS	unit.		
PS	Disclosure: Figure 1: 69pp; English.		
CC	The presence of a group of fungi may be detected in a sample by		
CC	detecting all or part of a double stranded molecule which comprises a		
CC	sequence encoding the 5S rRNA gene or a non-transcribed spacer		
CC	region of the rRNA repeat unit. The method is useful for detecting		
CC	specific fungal infections by amplifying a conserved repetitive DNA		
CC	region and concurrent amplification of an adjacent species-specific		
CC	sequence. This sequence contains the 5S rRNA gene of Candida		
CC	albicans as well as the non-transcribed spacer unit.		
SQ	Sequence 1200 BP: 367 A; 249 C; 245 G; 339 T;		
	Query Match 12.6%; Score 34.4; DB 1; Length 1200;		
	Best Local Similarity 48.9%; Pred. No. 0.17;		
	Matches 86; Conservative 0; Mismatches 90; Indels 0; Gaps		
OY	79 TATCTGTTGAAGCTACAGAAAGCATTTTANTTCAAAAATTGGCATTTTGATTCCT 138		
DB	405 TTTTAAGTTGAATTTATATAAAAATATTTACTCCAAATTTTGTCCAAATTTTGACA 464		
OY	139 GAACATCGAAGCAAGAAGTGACTATGTGAGACCAAACCTTAGACCTCAAAAATGCT 198		
DB	465 AAAAGTAAAAATAGAACCTTCCAATTTTTTTTAAACAACTTCAATCTAACCAACCGGAA 524		
OY	199 GATGTTCTGTTGCTGAGTCTANTTCCTCCAGNAATGATGAACCTTCACTGNGCA 254		
DB	525 AGTGTTGGGGGGCTGTGTAAACACACTGGTGTAAATAAAGTCCAAATATTTGGAA 580		
RESULT 3			
XI3748 ID	XI3748 standard: DNA; 799 BP.		
AC XI3748:			
DT 19-MAR-1999	(first entry)		
DE Enterococcus faecalis genome contig SEQ ID NO:811.			
KW Enterococcus faecalis; contig; detection; Enterococcal infection;			
KM vaccine; attenuation; computer readable medium; ds.			
OS Enterococcus faecalis.			
PN MO9850555-A2.			
PD 12-NOV-1998.			
PF 04-MAY-1998; U08985.			
PR 14-NOV-1997; US-066009.			
PR 06-MAY-1997; US-044031.			
PR 16-MAY-1997; US-046535.			
PA (HUMA-) HUMAN GENOME SCI INC.			
PI Barash SC, Dillon PJ, Kunsch CA;			
DR WPI: 99-045171/04.			
PT New isolated Enterococcus faecalis polynucleotides and peptides			
PT used to develop products for the detection of Enterococcus and for			
PT use in vaccines for prevention or attenuation of Enterococcus			

[illegible]

```
DB 1008 CCAACTATTTCGTTGAAACGTTTAACTTCAACAGACATCGGAGATGTGACTTT 1067
QY 174 AAAAATTGAGACCTACAAAATCTGATGTTGGTCTGAGCTTAACTTCCAGGMAA 233
DB 1068 CATACTAACATCTTCCAGGAATCTGATGATATGATGAAGATTAACCTCCAGGACA 1127
QY 234 TGATGAACCTTACACTGNGCAGA 256
DB 1128 AGTAGATCCAGACAGAGTGGACA 1150

RESULT 5
X08972
ID X08972 standard; DNA; 1947 BP.
AC X08972;
DE 12-MAY-1999 (first entry)
DE Human herpesvirus-7 gene U14.
KW Human herpesvirus-7; U14; pp85; antigenic determinant; epitope;
KW Immunoresponse; serology; HHV-7; HHV-6; detection; diagnosis; ss.
OS Human herpesvirus.
FH Key . Location/Qualifiers
FT CDS 1.1947
FT /tag- a
FT /product- pp85_protein
FT PN MO9902554-A2.
PD 21-JAN-1999.
PF 10-JUL-1998; U14372.
PR 10-JUL-1997; US-891672.
PA (UYBO-) UNIV BOLOGNA.
PA (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.
PI CAMPADILLI-Fiume G, Secchiello P;
PI WPI: 99-120778/10.
DR P-PSDB: W6199.
PT Human herpesvirus-7 pp85 protein - used in the diagnosis of human
PT herpesvirus-7 infection
PS Claim 3: Page 36-37: 54pp; English.
CC The human herpesvirus-7 (HHV-7) U14 gene encodes protein pp85, the
CC major determinant of immunoresponse to HHV-7. The pp85 protein and
CC its peptides (described in W96201-W96204 and having no serological
CC cross reactivity with human herpesvirus 6) can be used in a method
CC for the serological detection of HHV-7 in a biological sample. They
CC can also be used for diagnosing HHV-7 infection.
SO Sequence 1947 BP; 683 A; 372 C; 302 G; 590 T;

Query Match 11.6%; Score 31.8; DB 1; Length 1947;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 72; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 114 CAAAAATGTTCCATTGTTGATTCCTGAAACATGGAAGACAAAGAGTACTATGTGAGACC 173
DB 1008 CCAACTATTTCGTTGAAACGTTTAACTTCAACAGACATCGGAGATGTGACTTT 1067
QY 174 AAAAATTGAGACCTACAAAATCTGATGTTGGTCTGAGCTTAACTTCCAGGMAA 233
DB 1068 CATACTAACATCTTCCAGGAATCTGATGATATGATGAAGATTAACCTCCAGGACA 1127
QY 234 TGATGAACCTTACACTGNGCAGA 256
DB 1128 AGTAGATCCAGACAGAGTGGACA 1150

RESULT 6
N91071
ID N91071 standard; DNA; 2980 BP.
AC N91071;
DE 17-JUL-1990 (first entry)
DE Nucleotide sequence including open reading frame (ORF) of the btl3 gene
KW Bacillus thuringiensis tenebrionis; crystal protein; btl3 gene;
KW Coleoptera; P1VE38.
OS Bacillus thuringiensis.
FH Key . Location/Qualifiers
FT CDS 566.2500
```

```
FT misc-feature 743..771
FT /tag- a
FT /tag- b
FT /note- "synthetic DNA probe used for screening cosmid clones"
FT misc-difference 743..
FT /tag- c
FT /note- "mismatch between btl3 gene and synthetic DNA probe. Probe has C."
FT misc-difference 745..
FT /tag- d
FT /note- "As above. Probe has inosine."
FT misc-difference 748..751
FT /tag- e
FT /note- "As above. Probe has inosine-GAT."
FT misc-difference 757
FT /tag- f
FT /note- "As above. Probe has C."
FT misc-difference 760
FT /tag- g
FT /note- "As above. Probe has inosine."
FT misc-difference 765
FT /tag- h
FT /note- "As above. Probe has inosine."
FT misc-feature 542..544
FT /note- "in-frame ATG-codon of the btl3 gene"
FT misc-feature 566..568
FT /note- "in-frame ATG-codon of the btl3 gene"
FT misc-feature 707..709
FT /note- "in-frame ATG-codon of the btl3 gene"
FT misc-feature 527..529
FT /note- "first in-frame upstream stop codon"
FT rbs 534..559
FT /note- "putative Shine-Dalgarno"

W08901515-A.
PD 23-FEB-1989.
PR 15-AUG-1988; E00752.
PR 29-DEC-1987; GB-030261.
PA (PLAN-) Plant Genetic Systems.
PA Vaeck M, Hofte H, Botterman J;
DR WPI: 89-068871/09.
DR P-PSDB: P94679.
PT Plant cell transformed with DNA -
PT is for Bacillus thuringiensis toxic protein and is resistant to
PT Coleoptera pests
PS Disclousure; 35pp; English.
CC The btl3 gene is identical in nucleotide sequence to: 1) the B.t. gene
CC that is disclosed in European patent publication 0,213,818 (1987); and
CC the gene found in the B.t. strain which is disclosed in European patent
CC publication 0,149,162 (1985). The patent is for a transformed
CC plant cell which includes, inserted in its genome, the btl3 gene, or a
CC fragment which encodes a 66KD protein. The btl3 fragment is residues
CC 737-2497. The 66 kD protein is the active component of toxic crystal
CC protein produced by Bacillus thuringiensis and active against
CC Coleoptera. Plant cells contg. DNA encoding this protein are
CC protected against Coleoptera. A typical cassette for expression
CC in a plant contains the btl3 gene fused to a neo gene fragment (encoding
CC active neomycin phosphotransferase enzyme II), attached to the 35S
CC 3 promoter (5'-end) and to the polyadenylation/termination signals
CC of T-DNA gene 7 (3'-end). The prod. was P1VE38.
SO Sequence 2980 BP; 1095 A; 466 C; 513 G; 906 T;

Query Match 11.6%; Score 31.6; DB 1; Length 2980;
Best Local Similarity 51.4%; Pred. No. 1.8;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 GTTGCATGACCCCATCTGCCAGAGATGAACACTATTCACAAATTAAGCATG 60
DB 2524 GTGACCATCTATGATGTAGCAAGATTAATAATGATTAATAATACATA 2583
QY 61 GTGACCCAGCAATCCTGATCTGTTGAAGCTACAGAAACGATTTATTTCAAAAT 120
```

```
Db 2584 GTGTTCTCAACTTCGCTTTTGAGTAGATGAGAACACTATTTTATTTCAAAAT 2643
QY 121 GTGGCATTGATTCCTGAAA 142
Db 2644 GAAGGAAGTTTAAATATGTAA 2665

RESULT 7
N81002
ID N81002 standard: DNA: 2615 BP.
AC N81002:
DE 03-NOV-1990 (first entry)
KW Bacillus thuringiensis var. tenebrionis toxin gene;
KW Coleopteran insects; transformed plants; chimeric gene; ss.
OS Bacillus thuringiensis.
FH Key Location/Qualifiers
FT cds 205..2136
   /*tag= a
   /product=B.t.t. insecticidal toxin.
EP-289479-A.
PD 02-NOV-1988:
PF 26-APR-1988: 870070.
PR 29-APR-1987: US-044081.
PA (MONS) Monsanto CO.
PI Fischhoff DA, Fuchs RL, McPherson SA, Lavrik PB, Perlak FJ;
DR WPI: 88-309416/44.
P-PSDB: P80467.
DE Genetically transformed plants with toxicity to Coleopteran insects
PT - obtd. using chimeric gene contg. sequence encoding toxin protein
PS of Bacillus thuringiensis.
DE Disclosure: P: English.
CC A total of 2615 bp were obtained from pMON5420 and pMON5421.
CC The toxin gene is 1932 bp long encoding a protein of 644 amino acids
CC with mol.wt. of 73 091 kD. The protein has a net charge of -17 and
CC a G-C content of 34%.
CC See also N81000-N81003.
SQ Sequence 2615 BP; 952 A; 407 C; 473 G; 783 T;

Query Match
Best Local Similarity 11.6%; Score 31.6; DB 1; Length 2615;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 GTTGCATGACCCCAATGTGCCAGAGATGAACACTCATTCAACAATAAGACATG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2163 GTGACCATCTATGATAGTAGAAGCAAGATATAAAATGATTCATAATGAATACATA 2222
QY 61 GTGACCCAGGCACTCTGTATCTGTTTGAAGCTACAGCAAGAGATTTTCAAAAAT 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2223 GTGTTTTCACACTTCGCTTTTGAAGTAGATGAAGAACACTATTTTATTTCAAAAT 2282
QY 121 GTTGCATTTTGATTCCTGAAA 142
Db 2283 GAAGGAAGTTTAAATATGTAA 2304

RESULT 8
N82139
ID N82139 standard: DNA: 2980 BP.
AC N82139:
DE 01-NOV-1990 (first entry)
DE Cryc gene.
KW Coleoptera; toxin; cryc; insecticide; delta-endotoxin; Lepidoptera; ss.
OS Bacillus thuringiensis Bg2158.
FH Key Location/Qualifiers
FT cds 569..2503
   /*tag= a
   /product=Cryc protein
FT signal_peptide 569..727
   /*tag= b
FT mat_peptide 728..2500
   /*tag= c
   /product=cryc-toxin
```

```
FT rbs 557..563
   /*tag= d
PN MO8808880-A.
PD 17-NOV-1988:
PF 04-MAY-1987: U01495.
PR 06-MAY-1987: US-047945.
PA (ECOG-) Ecogen Inc.
PI Donovan WP, Gonzales JM, Levinson BL, Macaluso A;
DR WPI: 88-338229/47.
P-PSDB: P82487.
DE New gene encoding delta endotoxin of Bacillus thuringiensis -
PT effective as insecticide against Coleoptera.
PS Claim 1; Fig 8: 82pp; English.
CC The cryc gene encodes a delta endotoxin which is an insecticide
CC effective against Coleoptera and Lepidoptera. The sequence was
CC isolated from cDNA library, on a plasmid designated pBg212. The
CC plasmid can be used to construct an expression vector which can
CC be used to transform B. megaterium ATCC 35965. Colorado potato
CC beetle larvae treated with a suspension of cultured cells were
CC dead within one week. The gene may alternatively be inserted
CC directly into a plant which can then itself produce the active
CC toxin.
CC See also N82141.
SQ Sequence 2980 BP; 1097 A; 465 C; 514 G; 904 T;

Query Match
Best Local Similarity 11.6%; Score 31.6; DB 1; Length 2980;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 GTTGCATGACCCCAATGTGCCAGAGATGAACACTCATTCAACAATAAGACATG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2227 GTGACCATCTATGATAGTAGAAGCAAGATATAAAATGATTCATAATGAATACATA 2586
QY 61 GTGACCCAGGCACTCTGTATCTGTTTGAAGCTACAGCAAGAGATTTTCAAAAAT 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2587 GTGTTTTCACACTTCGCTTTTGAAGTAGATGAAGAACACTATTTTATTTCAAAAT 2646
QY 121 GTTGCATTTTGATTCCTGAAA 142
Db 2647 GAAGGAAGTTTAAATATGTAA 2668

RESULT 9
Q71026
ID Q71026 standard: DNA: 2983 BP.
AC Q71026:
DE 27-MAR-1995 (first entry)
DE CryIIIA gene which encodes insecticidal crystal protein.
KW CryIIIA; CryIIIA; CryIIB; CryC; P-2; CryBI; insecticidal protein crystal;
KW Lepidoptera; environmental insecticide; Bacillus thuringiensis; toxic;
KW probe; hybridisation; ss.
OS Bacillus thuringiensis.
FH Key Location/Qualifiers
FT rbs 557..562
   /*tag= a
   /product=CryIIIA protein
FT cds 569..2503
   /*tag= b
   /product=CryIIIA protein
PN US5338544-A.
PD 16-AUG-1994:
PF 16-APR-1987: 039542.
PR 16-APR-1987: US-039542.
PR 11-JUL-1989: US-379015.
PR 28-AUG-1991: US-751452.
PR 26-FEB-1993: US-023736.
PA (ECOG-) ECOGEN INC.
PI Donovan WP;
DR WPI: 94-263236/32.
P-PSDB: R56697.
DE New Cry IIB protein - obtd. from the cry II B gene in Bacillus
DE thuringiensis var. kurstaki, active against lepidopteran insects
FT Example 7; Fig 4A-4D; 39pp; English.
```



CC This sequence encodes CryIIIA crystal protein. The promoter from  
CC this sequence was fused to the protein coding region of the CryII  
CC gene (Q71077). The recombinant hybrid fusion gene expressed the CryII  
CC crystal protein more efficiently than its native promoter. Bacillus  
CC thuringiensis var. kurstaki produces crystal proteins during sporulation  
CC which are specifically toxic to certain orders and species of insects,  
CC esp. Lepidoptera. CryIIIB can be used in compositions used as insecticides,  
CC environmentally acceptable insecticides. (See also Q71025 and Q71027)  
SQ Sequence 2983 BP; 1097 A; 466 C; 514 G; 906 T;

Query Match 11.6%; Score 31.6; DB 1; Length 2983;  
Best Local Similarity 51.4%; Pred. No. 1.8;  
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

OY 1 GTTGCATGACCCCAATGTCGCAAGATGAACACATCATTCACAAATTAAGACATG 60  
DB 2527 GTGACCATGTATGATAGTAGAAGCAAGATAAATAAGATTATATAATGATACATTA 2586  
OY 61 GTGACCCAGCATCTCTGTATCTGTTGAAGCTACAGAAAGCATTTTATTTCAAAAT 120  
DB 2587 GTGCTTCTCACTTTCGCTTTTGAAGTAGATGAAGAACACTATTTTATTTCAAAAT 2646  
OY 121 GTTGCATTTTGTATTCCTGAAA 142  
DB 2647 GAAGGAAGTTTAAATATGTAA 2668

RESULT 10  
T33187  
ID T33187 standard; DNA: 2615 BP.  
AC T33187;

DT 04-DEC-1996 (first entry)  
DE Bacillus thuringiensis var. tenebrionis coleopteran toxin gene.  
KW Bacillus thuringiensis var. tenebrionis; insect resistance;  
KW transgenic plant; crop protection; crystal protein; toxin;  
KW Coleoptera; tomato; Lycopersicon esculentum; potato;  
KW Solanum tuberosum; corn; Zea mays; ds.  
OS Bacillus thuringiensis var. tenebrionis.  
FT Key Location/Qualifiers  
FT cds 205..2139  
FT /tag= a

PN EP-731170-A1.  
PD 11-SEP-1996.  
PR 26-APR-1988; 870070.  
PR 29-APR-1987; US-044081.  
PA (MONS.) MONSANTO CO.  
PI Fischhoff DA, Fuchs RL, Lavrik PB, McPherson SA;  
PI Perlak FJ;  
PI WPI: 96-403991/41.  
DR P-PSDB: R99958.  
PT Chimeric plant gene which expresses a Bacillus thuringiensis toxin  
PT protein - useful for prodn. of plants which are toxic to  
PT Coleopteran insects  
PS Disclosure: Fig 5A-J; 72pp; English.  
CC The gene (T33187) coding for the Coleopteran toxin (R99958) of  
CC Bacillus thuringiensis var. tenebrionis (B.t.t.) was obtd. by  
CC Southern hybridisation of B.t.t. total DNA using probes  
CC (T33188-86) based on the N-terminal portion (R99956) of  
CC the toxin. The gene can be incorporated into a vector, pref.  
CC under control of a plant-specific promoter, and utilised in  
CC the generation of Coleopteran insect-resistant transgenic  
CC plants, partic. tomato and corn.  
SQ Sequence 2615 BP; 949 A; 409 C; 475 G; 782 T;

Query Match 11.6%; Score 31.6; DB 1; Length 2615;  
Best Local Similarity 51.4%; Pred. No. 1.7;  
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

OY 1 GTTGCATGACCCCAATGTCGCAAGATGAACACATCATTCACAAATTAAGACATG 60  
DB 2163 GTGACCATGTATGATAGTAGAAGCAAGATAAATAAGATTATATAATGATACATTA 2222

OY 61 GTGACCCAGCATCTCTGTATCTGTTGAAGCTACAGAAAGCATTTATTTCAAAAT 120  
DB 2223 GTGCTTCTCACTTTCGCTTTTGAAGTAGATGAAGAACACTATTTTATTTCAAAAT 2282  
OY 121 GTTGCATTTTGTATTCCTGAAA 142  
DB 2283 GAAGGAAGTTTAAATATGTAA 2304

RESULT 11  
V21209\_01/c  
Continuation (2 of 17) of V21209 from base 100001 (Methanococcus jannaschii circular  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209  
WP Fragment Name Begin End  
WP V21209\_00 1 110000  
WP V21209\_01 100001 210000  
WP V21209\_02 200001 310000  
WP V21209\_03 300001 410000  
WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000  
WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000  
WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976

Query Match 11.4%; Score 31.2; DB 1; Length 110000;  
Best Local Similarity 52.0%; Pred. No. 9.2;  
Matches 66; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 113 TCAAAATGTTGCATTTGATTCCTGAAACATGAAAGCAAGNTGACTATGTGAGAC 172  
DB 77509 TAAATAGGTTAAGATAGAGAGCTCCCTATATGACAAAACAAAGAGGATGTAAAT 77450  
OY 173 CAAACTTGAGACCTTACAAAATGCTGATGTTCTGTTCTGATCTANTCTCCAGGAA 232  
DB 77449 TAGGGGCTGAATTTGGAAGAAAGCTTGTTGAGGTTTGAATAATAGTTATGATGCT 77390  
OY 233 ATGATGA 239  
DB 77389 ATCATGA 77383

RESULT 12  
V11696  
ID V11696 standard; DNA: 3475 BP.  
AC V11696;  
DT 14-SEP-1998 (first entry)  
DE Plasmid pCD34.  
KW Divergicin A; bacteriocin; signal peptide; protein secretion;  
KW preservative; food spoilage; lactic acid bacterium; antibacterial;  
KW plasmid pCD34; vector; ss; circular; cyclic.  
OS Carobacterium divergens strain LV13.  
PN W09809639-A1.  
PD 12-MAR-1998.  
PR 05-SEP-1997; U15609.  
PR 05-SEP-1996; US-026257.  
PI (UNL-) UNIV ALABAMA  
PI Franz CW, Greer GG, Lelsner JT, McCormick JR, McMullen LM,  
PI Pkns J, Poona, Stiles WE, Van Belkum MJ, Vederas JC,  
PI Worobo RJ, Worobo RW;  
PI WPI: 98-193319/17.  
DR Bacterial growth inhibiting peptide(s) enterocin 900 or brochozin C  
PT - used to inhibit growth of susceptible bacteria in e.g. foodstuff,  
PT living animal, food preparation area or fermentation vessel

PS Claim 14; Page 127-129; 174pp; English.  
 CC PC03.4 comprises a food-grade plasmid from *Carnobacterium divergens*  
 CC LV13, a producer of the class II bacteriocin divergicin A. The  
 CC plasmid can be used to insert genes for use in probiotics, meat,  
 CC milk products, food or food products. PC03.4 can be used in new  
 CC methods for exporting bacteriocins from cells using divergicin A  
 CC signal peptide (see W59007). The invention provides plasmid  
 CC vectors in which a signal peptide or leader peptide sequence is  
 CC attached to DNA encoding a bacteriocin and to an immunity gene  
 CC specific for the bacteriocin. Bacteriocin cassettes encoding the  
 CC divergicin A signal peptide and two or more bacteriocins are used to  
 CC tailor the antibacterial spectrum of a producer strain to target a  
 CC range of spoilage or pathogenic bacteria, e.g. *Escherichia coli* and  
 CC *Salmonella*. Using the technology of multiple bacteriocin production  
 CC and delivery using lactic acid bacteria, a range of bacteriocins can  
 CC be produced in situ in human food, animal feed, the gastrointestinal  
 CC tract of humans and animals, and in the female genital tract.  
 SO Sequence 3475 BP; 1209 A; 482 C; 605 G; 1179 T;

Query Match 11.1%; Score 30.4; DB 1; Length 3475;  
 Best Local Similarity 52.9%; Pred. No. 4.4;  
 Matches 64; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
 QY 90 AGCTACAGGAACGCGATTATTTTCAAAAATGTCGCTTTGATTCCTGGAACATGGAA 149  
 DB 372 AGTTTGAAGAGTTTATTTTATTTGATACCAATTTGCTTTTATTAATTAAGATATA 431  
 QY 150 GACAAAGNTGACTATGTGAGACCAAACTTGACACTACAAAATGCGATGTTCTGCT 209  
 DB 432 CAATATAGCGACTATTTTAAATTTACAACTTTTGAATTAAGTATATTTGTTATAT 491  
 QY 210 T 210  
 DB 492 T 492

RESULT 13

X17720  
 ID X17720 standard; DNA; 3475 BP.  
 AC X17720;  
 DT 25-MAY-1999 (first entry)  
 DE Food-grade plasmid PC03.4.  
 KM Enterocin 900; brochocin-C; bacterial growth; microorganism; inhibit;  
 KM food; meat; fermentation; bacteriocin; mastitis; secretory vector; lacc;  
 KM leucocin A; lacc; enzyme; probiotic; L. gelidium; *Listeria*; *Lactobacillus*;  
 KM sulphide-producing; spoilage; sensory property; smell; taste; infection;  
 KM ground beef; silage; *Salmonella*; poultry; intestine; therapeutic; colour;  
 KM mucosal tissue; vacuum-packed; storage; circular; ss.  
 OS Synthetic.  
 PN MO9902555-A1.  
 PD 21-JAN-1999.  
 PF 09-JUN-1998; U14331.  
 PR 05-SEP-1997; US-924629.  
 PR 09-JUL-1997; US-052072.  
 PA (UVAL-) UNIV ALBERTA.  
 PI Franz C, Greer GG, Leisner JJ, McCormick JK, McWilliam LM,  
 PI Poon A, Stiles ME, Van Belkum MJ, Vederas JC, Worobo RJ,  
 PI Morobo RW.  
 DR WPI: 99-12079/10.  
 PT New bacteriocins enterocin 900 and brochocin-C - useful as food  
 PT preservatives and therapeutic antimicrobials  
 PS Claim 14; Page 126-128; 182pp; English.  
 CC The invention relates to peptides enterocin 900, brochocin-C that inhibit  
 CC bacterial growth. Microorganisms expressing the above peptides are used  
 CC to inhibit bacterial growth in foods (specifically meat), live animals  
 CC (applied topically), food preparation areas and fermentation vessels. The  
 CC bacteriocins enterocin 900, brochocin-C may be used directly, specifically  
 CC to inhibit bacteria that cause mastitis. Antibodies specific for the  
 CC peptides are used to detect expression of bacteriocins in cells. Secretory  
 CC vectors (containing sequences for a promoter, a leucocin A processing  
 CC peptide, a heterologous peptide, lacc and lacc) are used to process  
 CC bacteriocins, enzymes or other proteins, e.g. for use in food production

CC and as probiotics. *L. gelidium*, or other leucocin A-expressing bacteria,  
 CC inhibit the growth of *Listeria* and the sulphide-producing spoilage  
 CC organism *Lactobacillus sake*; they also improve the sensory properties  
 CC (smell, taste and colour) of meat, specifically ground beef. *L. gelidium*  
 CC can also be used in preparation of animal feeds (silage), as probiotic,  
 CC to control *Salmonella* in poultry intestines and therapeutically against  
 CC mucosal tissue infections. Treatment with bacteriocins provides vacuum-  
 CC packed meat products with predictable and longer storage life. The  
 CC secretion vectors can express several bacteriocins for broader spectrum  
 CC of activity, e.g. tailored for a particular target organism. The present  
 CC sequence represents a claimed plasmid vector PC03.4, which is a  
 CC food-grade plasmid.  
 SO Sequence 3475 BP; 1209 A; 481 C; 606 G; 1179 T;

Query Match 11.1%; Score 30.4; DB 1; Length 3475;  
 Best Local Similarity 52.9%; Pred. No. 4.4;  
 Matches 64; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
 QY 90 AGCTACAGGAACGCGATTATTTTCAAAAATGTCGCTTTGATTCCTGGAACATGGAA 149  
 DB 372 AGTTTGAAGAGTTTATTTTATTTGATACCAATTTGCTTTTATTAATTAAGATATA 431  
 QY 150 GACAAAGNTGACTATGTGAGACCAAACTTGACACTACAAAATGCGATGTTCTGCT 209  
 DB 432 CAATATAGCGACTATTTTAAATTTACAACTTTTGAATTAAGTATATTTGTTATAT 491  
 QY 210 T 210  
 DB 492 T 492

RESULT 14

V52150/c  
 ID V52150 standard; DNA; 9064 BP.  
 AC V52150;  
 DT 23-OCT-1998 (first entry)  
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:17.  
 KM Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KM computer readable medium; vaccine; pharmaceutical composition; ds.  
 OS Streptococcus pneumoniae.  
 PN MO9818931-A2.  
 PD 07-MAY-1998.  
 PF 30-OCT-1997; U19588.  
 PF 31-OCT-1996; US-029960.  
 PA (HMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
 PI Kunsch CA, Rosen CA;  
 DR WPI: 98-27225/24.  
 PT Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 PS Claim 1; Page 244-249; 1409pp; English.  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S. pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridise to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid molecules derived from the  
 CC sequence is homologous to amplification primers the amplification and  
 CC fragment of the S. pneumoniae genome to prime the amplification can be  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating

CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
SQ Sequence 9064 BP; 3049 A; 1825 C; 1326 G; 2864 T;

Query Match 10.9%; Score 29.8; DB 1; Length 9064;

Best Local Similarity 50.7%; Pred. No. 9.7;  
Matches 70; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

OY 69 GGCATCTCTGATCTGTGACCTACAGGAAGCGATTTTATTCAAAATGTGCAT 128  
DB 6496 GTCCCTTTGTGATGATGATCTTAATATGATATTTATTCCTATTAATTTACTGT 6437  
OY 129 TTGATTCCTGAAACATGGAAGCAAGGNTGACTATGTGAGACCAAACTTGAGACTA 188  
DB 6436 CTTATTTACTTACTTACAAATATCAAGTTATCTAAGATGATGAGAAATAGATAGTTA 6377  
OY 189 CAAAATGCTGATGTTCT 206  
DB 6376 GCAAAATTTTAATATAT 6359

## RESULT 15

V21209\_15/c

Continuation (16 of 17) of V21209 from base 1500001 (Methanococcus jannaschii circular  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 10.8%; Score 29.6; DB 1; Length 110000;

Best Local Similarity 55.4%; Pred. No. 28;  
Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 106 TTTTATTTCAAAAATGTGCAATTTGATTCCTGAAACATGGAAGCAAGGNTGACTAT 165  
DB 33907 TTAGATATAGAAAAATGCAATTTGAAATGCTGAAAGCTACTATGAAAAAGCTAAAAAG 33848  
OY 166 GTGAGACCAAACTTGAGACTTACAAAAAATGCTGATGTTCT 206  
DB 33847 TTGAGAAATATAAATGAAAGATAGAAATGCTATGAGCT 33807

Search completed: August 6, 1999, 04:01:48  
Job time: 1460 sec

**This Page Blank (uspto)**



```

PF 14-MAR-1991: U01748.
PR 14-MAR-1990: US-494036.
PA (WASH-) WASHINGTON RES FOUN.
PI Fischer EH, Krebs EG, Tonks NK, Cool DE.
PI WPI: 91-295643/40.
DR P-PSDB: R14114.
PT New DNA encoding non-receptor-linked protein tyrosine phosphatase
PT used in treating malignancies associated with protein-linked
PT tyrosine phosphatase.
PS Claim 1; Fig 1; 38pp; English.
CC A CDNA library prep. from human periphera T cell poly(A)+ mRNA was
CC screened using probes designed from the sequence of human placenta
CC prase 1b. The sequence can be used to express recombinant pPase
CC or derivs. useful in cancer therapy where a protein kinase is
CC involved.
CC See also Q13817-Q13820.
CC Sequence 2287 BP; 753 A; 411 C; 464 G; 659 T;
SQ

Query Match 6.2%; Score 17; DB 1; Length 2287;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CATCTGTGATCTGTTT 87
DB 211 CATCTGTGATCTGTTT 195

RESULT 3
V52044
ID V52044 standard; DNA: 1883 BP.
AC V52044;
DE 09-NOV-1998 (first entry)
DE Helicobacter polypeptide GHPD 1284 DNA.
KM GHPD 1284; infection; therapy; diagnosis; vaccine; gastritis;
KM ulcer; ss.
OS Helicobacter pylori.
FH Key
FT Location/Qualifiers
FT CDS 91..1836
FT /tag= a
FT
PD M09821225-A1.
PD 22-MAR-1998.
PR 14-NOV-1997; U21353.
PR 29-JUL-1997; US-902615.
PR 14-NOV-1996; US-749051.
PR 01-APR-1997; US-831309.
PR 01-APR-1997; US-833457.
PR 01-APR-1997; US-834705.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (PLNC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Haas R, Kleanthous H, Meyer T, Miller C,
PI Odenbreit S, Tomb J;
PI WPI: 98-297855/26.
DR P-PSDB: W71509.
PT Helicobacter polynucleotide and polypeptide sequences - useful to
PT treat or prevent gastrointestinal infection
PS Disclosure: Page 176-179; 362pp; English.
CC This polynucleotide encodes Helicobacter pylori polypeptide
CC GHPD 1284 (see W71509). 85 polynucleotides (see V52009-V52093)
CC encoding claimed Helicobacter polypeptides (see W71474-W71558) have
CC been identified in the H. pylori genome using a transposon shuttle
CC mutagenesis method in which Tmex9 mini-plam was used for
CC insertional mutagenesis of an H. pylori gene library established in
CC Escherichia coli. The invention provides: methods for producing
CC these polypeptides in recombinant host systems, and related
CC expression cassettes, vectors and transformed or transfected host
CC cells; live vaccine vectors that contain polynucleotides of the
CC invention and which can be used to prevent or treat Helicobacter
CC infection; therapeutic and/or prophylactic methods involving
CC administration of polynucleotide molecules (either in naked form
CC or delivered with a delivery vehicle), polypeptides or monospecific

```

```

CC antibodies; methods for detecting the presence of Helicobacter
CC in samples using e.g. polynucleotide molecules; and methods for
CC purifying polypeptides of the invention.
SQ Sequence 1883 BP; 659 A; 296 C; 415 G; 512 T;

Query Match 6.2%; Score 17; DB 1; Length 1883;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ATGTGCCAGAGATGAA 33
DB 953 ATGTGCCAGAGATGAA 969

RESULT 4
X13194/C
ID X13194 standard; DNA: 7992 BP.
AC X13194;
DE 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:257.
KM Enterococcus faecalis; contig; detection; Enterococcal infection;
KM vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN M09850555-A2.
PD 12-NOV-1998.
PR 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
PI WPI: 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT used to develop products for the detection of Enterococcus
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 1235-1239; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 992 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
CC Sequence 7992 BP; 2405 A; 1576 C; 1049 G; 2943 T;
SQ

Query Match 6.2%; Score 17; DB 1; Length 7992;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CAAAATGCTATGTTTC 205
DB 6287 CAAAATGCTATGTTTC 6271

RESULT 5
X14537/C
ID X14537 standard; DNA: 1025 BP.
AC X14537;
DE 31-MAR-1999 (first entry)
DE H. pylori GHPD 1408 gene.
KM GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
KM peptic ulcer disease; ss.
OS Helicobacter pylori.

```

FH Key Location/Qualifiers  
 FT CDS 22..1002  
 FT CDS /tag- a  
 PN WO9843478-A1.  
 PD 08-OCT-1998.  
 PF 01-APR-1998; US-902615.  
 PR 29-JUL-1997; US-902615.  
 PR 01-APR-1997; US-833457.  
 PR 24-JUN-1997; US-881227.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 PI Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;  
 DR WPI: 98-542293/46.  
 DR P-PSDB; W98818.  
 PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 PS Claim 1: Page 1798-1799; 2054pp; English.  
 CC This sequence represents a polynucleotide of the invention. It was  
 CC isolated from Helicobacter pylori and encodes a H.pylori GHPD protein.  
 CC The polypeptides can be used for preventing or treating Helicobacter  
 CC infections, and gastroduodenal diseases associated with these  
 CC infections, including acute, chronic, and atrophic gastritis, and peptic  
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used  
 CC for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.  
 SQ Sequence 1025 BP; 323 A; 157 C; 294 G; 251 T;  
 Query Match 6.2%; Score 17; DB 1; Length 1025;  
 Best Local Similarity 100.0%; Pred. No. 6.8;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 122 TTGCATTGATTCCT 138  
 DB 28 TTGCCATTGATTCCT 12  
 RESULT 6  
 ID X14246 standard; DNA; 1198 BP.  
 AC X14246:  
 DT 31-MAR-1999 (first entry)  
 DE H. pylori GHPD 1407 gene.  
 KW GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease; ss.  
 OS Helicobacter pylori.  
 FH Key Location/Qualifiers  
 FT CDS 55..1148  
 FT CDS /tag- a  
 PN WO9843478-A1.  
 PD 08-OCT-1998.  
 PF 01-APR-1998; US-902615.  
 PR 29-JUL-1997; US-902615.  
 PR 01-APR-1997; US-833457.  
 PR 24-JUN-1997; US-881227.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 PI Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;  
 DR WPI: 98-542293/46.  
 DR P-PSDB; W98527.  
 PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 PS Claim 1: Page 1003-1005; 2054pp; English.  
 CC This sequence represents a polynucleotide of the invention. It was  
 CC isolated from Helicobacter pylori and encodes a H.pylori GHPD protein.  
 CC The polypeptides can be used for preventing or treating Helicobacter  
 CC infections, and gastroduodenal diseases associated with these  
 CC infections, including acute, chronic, and atrophic gastritis, and peptic  
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used  
 CC for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.

SQ Sequence 1198 BP; 352 A; 208 C; 327 G; 311 T;  
 Query Match 6.2%; Score 17; DB 1; Length 1198;  
 Best Local Similarity 100.0%; Pred. No. 6.8;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 122 TTGCATTGATTCCT 138  
 DB 1176 TTGCCATTGATTCCT 1160  
 RESULT 7  
 ID 050871/c  
 AC 050871:  
 DT 13-MAY-1994 (first entry)  
 DE BSP 14-3-3  
 KW Brain specific protein; BSP, plant cell; callus; activator;  
 KW thymosine monooxygenase; tryptophan monooxygenase; vector; ss.  
 OS Oryza sativa.  
 FH Key Location/Qualifiers  
 FT CDS 243..1022  
 FT CDS /tag- a  
 FT CDS /product- BSP14-3-3-protein-equivalent  
 PN J05260975-A.  
 PD 12-OCT-1993.  
 PF 23-MAR-1992; 065299.  
 PR 23-MAR-1992; JP-065299.  
 PA (POK) POEA CHEM IND INC.  
 DR WPI: 93-356451/45.  
 DR P-PSDB; R42198.  
 PT New plant gene with specified aminoacid sequence - codes protein  
 PT having functions equiv. to those of brain specific portion  
 PS Disclosure: Page 6-7; 12pp; Japanese.  
 CC The sequence encodes a plant protein having functions specific to  
 CC that of brain specific protein BSP14-3-3. The expression of this  
 CC gene in cultured plants cells, callus etc. brings forth activators  
 CC of enzymes e.g. thymosine monooxygenase and tryptophan  
 CC monooxygenase.  
 SQ Sequence 1465 BP; 366 A; 332 C; 387 G; 380 T;  
 Query Match 5.9%; Score 16; DB 1; Length 1465;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 45 ACAATRAAGACATG 60  
 DB 1340 ACAATRAAGACATG 1325  
 RESULT 8  
 ID T48019/c  
 AC T48019:  
 DT 29-MAY-1997 (first entry)  
 DE Encodes FIV principal immunodominant domain mutant peptide n14.  
 KW Feline immunodeficiency virus; principal immunodominant domain;  
 KW mutant; lentivirus; envelope protein; neutralising antibody;  
 KW vaccine; ss.  
 OS Feline immunodeficiency virus.  
 PN WO9630527-A1.  
 PD 03-OCT-1996.  
 PF 26-MAR-1996; F00449.  
 PR 27-MAR-1995; FR-003566.  
 PA (CNRS) CNRS CENT NAT RECH SCI.  
 PI Panchino G, Sonigo P;  
 DR WPI: 96-455372/45.  
 DR P-PSDB; W09332.  
 PT Mutant lentivirus Env protein principal immuno-dominant domains -  
 PT can produce neutralising antibodies to protect against e.g. FIV or  
 PT HIV, but do not produce facilitating antibodies

PS Claim 6, Page 21: 38bp: French.  
 CC The present sequence encodes a specifically claimed example of a novel  
 CC peptide fragment which is a mutant of the wild-type (w.t.) principal  
 CC immunodominant domain (PID) of feline immunodeficiency virus such  
 CC that an Env protein containing the fragment retains the ability to  
 CC produce neutralising antibodies (Abs) but does not produce  
 CC facilitating or deleterious Abs against the w.t. PID. The mutated  
 CC Env protein can be used for producing vaccines against FIV. In  
 CC addition to vaccination, the peptide fragment is useful as a  
 CC diagnostic reagent, specifically to distinguish between Abs produced  
 CC in response to vaccination and those resulting from viral infection.  
 CC Also, the results of anti-FIV vaccination can be monitored  
 CC using the peptide.  
 CC Sequence 27 BP: 8 A: 5 C: 4 G: 10 T:

Query Match  
 Best Local Similarity 100.0%; Score 16; DB 1; Length 27;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 190 AAAATGCTGATGTC 205  
 DB 19 AAAATGCTGATGTC 4

RESULT 9  
 V04698  
 ID V04698 standard; cDNA: 1507 BP.  
 AC V04698:  
 DT 21-JUL-1998 (first entry)  
 DE Homo sapiens 20q13 amplicon cc49 transcript.  
 CC 20q13 amplicon; chromosome 20: tumour; detection: C2H2 zinc finger;  
 CC chromosomal abnormalities; probe: gene therapy; antisense inhibition;  
 CC treatment; age-related macular degeneration; retinitis pigmentation;  
 CC Leber's congenital amaurosis; ds.  
 CC Homo sapiens.  
 CC WO9802539-A1.  
 PD 22-JAN-1998:  
 PF 15-JUL-1997; U12343  
 PR 17-JAN-1997; US-785332.  
 PR 15-JUL-1996; US-680395.  
 PR 16-OCT-1996; US-731499.  
 PA (RECC) UNIV CALIFORNIA.  
 PI Collins CC, Godfrey T, Gray JW, Hwang SI, Kowbel D,  
 PI Rommens J;  
 DR WPI: 98-110587/10.  
 PT New sequences from the 20q13 amplicon - used for detecting  
 PT chromosomal abnormalities, particularly tumours, and for developing  
 PT products for treating diseases  
 PS Claim 1: Pages 61-62; 91pp: English.  
 CC The sequence is that of a cDNA sequence cc49, which was isolated  
 CC from the 20q13 amplicon. It shows homology to C2H2 zinc finger  
 CC genes. It can be used as a probe for the detection  
 CC of chromosomal abnormalities at 20q13. It and other  
 CC sequences isolated from the 20q13 amplicon are consistently  
 CC amplified in primary tumours. These sequences are useful as probes  
 CC or as probe targets for monitoring the relative copy number of  
 CC corresponding sequences from a biological sample such as tumour  
 CC cells. The sequences can also be used in therapeutic applications  
 CC for modulating the expression of the endogenous gene or the activity  
 CC of the gene product. Examples of therapeutic approaches include  
 CC antisense inhibition of gene expression, gene therapy, and monoclonal  
 CC antibodies that specifically bind the gene products. The products can  
 CC also be used in the treatment of other diseases, e.g. age-related  
 CC macular degeneration, Leber's congenital amaurosis and retinitis  
 CC pigmentations.  
 CC Sequence 1507 BP: 445 A: 354 C: 364 G: 340 T:

OY 30 TGAACACTCATTCAA 45  
 DB 32 TGAACACTCATTCAA 47

RESULT 10  
 V09023  
 ID V09023 standard; DNA: 10282 BP.  
 AC V09023:  
 DT 21-JUL-1998 (first entry)  
 DE Homo sapiens 20q13 amplicon ZABC-1 genomic sequence.  
 CC 20q13 amplicon; chromosome 20: tumour; detection: ZABC-1 gene;  
 CC chromosomal abnormalities; probe: gene therapy; antisense inhibition;  
 CC treatment; age-related macular degeneration; retinitis pigmentation;  
 CC Leber's congenital amaurosis; zinc finger amplified in breast cancer;  
 CC ds.  
 CC Homo sapiens.  
 CC WO9802539-A1.  
 PD 22-JAN-1998:  
 PF 15-JUL-1997; U12343  
 PR 17-JAN-1997; US-785332.  
 PR 15-JUL-1996; US-680395.  
 PR 16-OCT-1996; US-731499.  
 PA (RECC) UNIV CALIFORNIA.  
 PI Collins CC, Godfrey T, Gray JW, Hwang SI, Kowbel D,  
 PI Rommens J;  
 DR WPI: 98-110587/10.  
 DR P-PSDB: W23975.  
 PT New sequences from the 20q13 amplicon - used for detecting  
 PT chromosomal abnormalities, particularly tumours, and for developing  
 PT products for treating diseases  
 PS Claim 1: Pages 64-67; 91pp: English.  
 CC The sequence is that of the genomic sequence of ZABC-1 (zinc finger  
 CC amplified in breast cancer). It maps to the core of the 20q13.2  
 CC amplicon and is overexpressed in primary tumours and breast cancer  
 CC cell lines having 20q13.2 amplification. The exact coding region for  
 CC the genomic sequence is not given. The sequence can be used as a  
 CC probe for the detection of chromosomal abnormalities at 20q13. It and  
 CC other sequences isolated from the 20q13 amplicon are consistently  
 CC amplified in primary tumours. These sequences are useful as probes  
 CC or as probe targets for monitoring the relative copy number of  
 CC corresponding sequences from a biological sample such as tumour  
 CC cells. The sequences can also be used in therapeutic applications  
 CC for modulating the expression of the endogenous gene or the activity  
 CC of the gene product. Examples of therapeutic approaches include  
 CC antisense inhibition of gene expression, gene therapy, and monoclonal  
 CC antibodies that specifically bind the gene products. The products can  
 CC also be used in the treatment of other diseases, e.g. age-related  
 CC macular degeneration, Leber's congenital amaurosis and retinitis  
 CC pigmentations.  
 CC Sequence 10282 BP: 2820 A: 2222 C: 2191 G: 3047 T:

Query Match  
 Best Local Similarity 100.0%; Score 16; DB 1; Length 10282;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 TGAACACTCATTCAA 45  
 DB 2453 TGAACACTCATTCAA 2468

RESULT 11  
 V21209\_01  
 Continuation (2 of 17) of V21209 from base 100001 (Methanococcus jannaschii circular  
 WP Sequence split into 17 fragments LOCUS V21209 Accession V21209  
 WP Fragment Name Begin End  
 WP V21209\_00 1 110000  
 WP V21209\_01 100001 210000  
 WP V21209\_02 200001 310000  
 WP V21209\_03 300001 410000  
 WP V21209\_04 400001 510000  
 WP V21209\_05 500001 610000





RESULT 15

X20848 standard; DNA; 417 BP.

AC X20848;

DT 05-MAY-1999 (first entry)  
polynucleotide sequence from the genome of *Treponema pallidum*.  
DE polynucleotide sequence from the genome of *Treponema pallidum*. an

KW Treponema pallidum infection; syphilis; Borrelia infection, bacterial  
KW enzyme production; ds.

OS Treponema pallidum.  
PN W09859034-A2.

PN MO3025024 AZ:  
PD 30-DEC-1998.  
03 TTY-1008.

PF 23-JUN-1998; 013041  
PR 24-JUN-1997; US-0506

PA (HUMA-) HUMAN GENOME  
PI Fraser CM;

DR WPI; 99-081273/07.  
PT New isolated *Treponema pallidum*.

new products for the detection, diagnosis, and therapy of T cell lymphoma in

Claim 1; F

CC X20500-21243 repr  
CC Treponema pallidum  
CC

diagnosis, characterisation, prevention and control of infectious diseases. They can also be used for the detection of antibodies in animals and for the diagnosis of infectious diseases.

CC diseases related to *Borrelia* infections in animals, and for the  
CC production of biosynthetic products such as enzymes.

Sequence 417 BP; 139 A; 51 C; 52 G; 168 T;

5 g: Score 16: DB 1: Length 41

Query Match	100.0%	Pred. No. 22;	0: Indels
Best Local Similarity	100.0%	Mismatches	0: Indels

Matches	16;	Conservative	0;	Mismatches	0;	Inserts	0;	Deletions	0;
---------	-----	--------------	----	------------	----	---------	----	-----------	----

QY 105 ATTTTATTTCAAAAT 120  
|||||

Db 66 ATTTATTCAAAAT 51

Search completed: August 5, 1999, 17:39:27  
Job time: 6202 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: August 5, 1999, 17:39:47 ; Search time 650.92 Seconds

(without alignments)  
69,570 Million cell updates/sec

Title: US-09-049-696-4

Perfect score: 181

Sequence: 1 CAAAAGATGCACATTCATA.....ACACGCAAAATCAAAATATGC 181

Scoring table: OLIGO\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	9.9	716	1	V68834
2	17	9.4	376	1	V68909
3	16	8.8	4104	1	V52302
4	16	8.8	2130	1	V43006
5	16	8.8	529	1	T98762
6	16	8.8	829	1	V59890
7	16	8.8	462	1	V75326
8	16	8.8	6076	1	V74617
9	16	8.8	528	1	X30886
10	15	8.3	12284	1	X06001
11	15	8.3	2642	1	O10350
12	15	8.3	2633	1	O10351
13	15	8.3	1153	1	O11661
14	15	8.3	1069	1	O11662
15	15	8.3	1498	1	O43504
16	15	8.3	1497	1	O41000
17	15	8.3	840	1	O51348
18	15	8.3	1860	1	O50632
19	15	8.3	12284	1	O71308
20	15	8.3	2556	1	T03197
21	15	8.3	2403	1	T11081
22	15	8.3	129	1	T20332
23	15	8.3	12284	1	T29591
24	15	8.3	1812	1	T27105
25	15	8.3	1837	1	T33852
26	15	8.3	1498	1	T38359
27	15	8.3	8603	1	T85473
28	15	8.3	8367	1	T73721
29	15	8.3	6803	1	V16305
30	15	8.3	15229	1	V18276
31	15	8.3	1642	1	V40507
32	15	8.3	110000	1	V21209_01
33	15	8.3	7705	1	T75298
34	15	8.3	7091	1	V69428
35	15	8.3	4291	1	V69278
36	15	8.3	15598	1	V74393
37	15	8.3	2670	1	V72215
38	15	8.3	2880	1	X18193
39	15	8.3	3912	1	X18192
40	15	8.3	35515	1	X20252
41	15	8.3	1514	1	X22018
42	15	8.3	505	1	X27373
43	15	8.3	360	1	X40634

## ALIGNMENTS

RESULT 1  
V68834  
ID V68834 standard; DNA; 716 BP.  
AC V68834;  
DT 22-JAN-1999 (first entry)  
DE DNA molecule encoding a breast tumour specific polypeptide #26.  
KW Human; breast cancer; breast tumour tissue; diagnosis; treatment;  
OS vaccine; epitope; endogenous; retroviral element; ss.  
PN Homo sapiens.  
PD WO9845328-A2.  
PF 15-OCT-1998.  
PR 09-APR-1998; U06939.  
PR 11-DEC-1997; US-891789.  
PR 09-APR-1997; US-838762.  
PI (CORI-) CORIXA CORP.  
PI Frudakis TM, Reed SG, Smith JM;  
DR WPI: 98-557473/47.  
PT New DNA sequences isolated from endogenous human retroviral element  
PT - and related vectors, transformed cells, proteins and antibodies,  
PT useful for diagnosis, treatment and prevention of breast cancer  
PS Claim 1; Page 49; 173pp; English.  
CC V68800 to V68998 represent nucleotide sequences which encode human  
CC breast tumour specific polypeptides. Detection or measurement of  
CC human breast tumour specific polypeptides and nucleotide sequences,  
CC or the corresponding RNA in a sample, is used for diagnosis and  
CC monitoring of breast cancer. Human breast tumour specific polypeptides  
CC and nucleotide sequences, and the vectors containing the DNAs, are also  
CC useful in vaccines for inhibiting development (for prevention or  
CC therapy) of breast cancer. The polypeptides may also be used to  
CC raise monoclonal antibodies, used as immunoassay reagents.  
SQ Sequence 716 BP; 244 A; 119 C; 121 G; 226 T;

Query Match 9.9%; Score 18; DB 1; Length 716;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 14 TTCATTAAGTAAACAGGA 31  
DB 266 TTCATTAAGTAAACAGGA 283  
RESULT 2  
V68909/c  
ID V68909 standard; cDNA; 376 BP.  
AC V68909;  
DT 27-APR-1999 (first entry)  
DE EST clone B488.  
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
KW tissue growth; actinin; inhibin; tumour invasion suppressor; EST; human;  
KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;  
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
OS Homo sapiens.  
PN WO9845435-A2.  
PD 15-OCT-1998.  
PR 10-APR-1998; U06954.  
PR 10-APR-1997; US-835913.  
PA (GENEY) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavalie ER, McCoy JM, Merberg D,  
PI Racle IA, Spaulding V, Treacy M;  
DR WPI: 99-070076/06.  
PT New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries  
PS Claim 1; Page 388; 633pp; English.  
CC This sequence represents an expressed sequence tag (EST), and is a

CC polynucleotide of the invention. The polynucleotides of the invention are  
CC all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC activity, thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC therapy.  
CC Sequence 376 BP; 101 A; 58 C; 71 G; 146 T;

Query Match 9.4%; Score 17; DB 1; Length 376;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 135 AACAAACACACACAA 151  
|  
Db 198 AACAAACACACACAA 182

RESULT 3  
V52302/c  
ID V52302 standard; DNA; 4104 BP.  
AC V52302;  
DT 23-OCT-1998 (first entry)  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:169.  
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.  
PN Streptococcus pneumoniae.  
PN MO9818931-A2.  
PD 07-MAY-1998.  
PF 30-OCT-1997; U19588.  
PR 31-OCT-1996; US-029960.  
PA (HUMAN) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
PI Kunsch CA, Rosen CA;  
PI WPI: 98-272225/24.  
DR Computer readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
PS Claim 1: Page 1078-1080; 1409pp; English.  
CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
CC on it, or a representative fragment of a sequence at least 95% identical  
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
CC to 391, identifying members of the library which contain sequences  
CC that hybridise to the target sequence and isolating the nucleic acid  
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
CC from an organism, amplifying nucleic acid molecules whose nucleotide  
CC sequence is homologous to amplification primers derived from the  
CC fragment of the S. pneumoniae genome to prime the amplification and  
CC isolating the amplified sequences. The computer readable medium can be  
CC used in a computer-based system for identifying fragments of the  
CC S. pneumoniae genome of commercial importance, or expression modulating  
CC fragments of the S. pneumoniae genome. Products from the present  
CC invention can be used in diagnosis kits and assays, and pharmaceutical  
CC compositions and vaccines for S. pneumoniae.  
CC Sequence 4104 BP; 1264 A; 711 C; 924 G; 1205 T;

Query Match

8.8%; Score 16; DB 1; Length 4104;

Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AACACAGCAAAATCA 175  
|  
Db 3185 AACACAGCAAAATCA 3170

RESULT 4  
V43006/c  
ID V43006 standard; DNA; 2130 BP.  
AC V43006;  
DT 09-NOV-1998 (first entry)  
DE Streptococcus pneumoniae polypeptide coding region.  
KW Polypeptide; ORF; open reading frame; infection; bacterial;  
KW Streptococcal; bacteremia; diagnosis; prophylaxis; ds.  
OS Streptococcus pneumoniae.  
FH Key Location/Qualifiers  
FT CDS 1915..2097  
FT /tag- a  
FT /note= "polypeptide"  
PN MO9823631-A1.  
PD 04-JUN-1998.  
PF 24-NOV-1997; U21976.  
PR 27-NOV-1996; US-031879.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Black MT, Hodgson JE, Knowles DCC, Lonetto MA, Nicholas RO,  
PI Reid RH, Zarfos PN;  
PI WPI: 98-322654/28.  
DR P-PDB: W62731.  
PT Streptococcus pneumoniae polynucleotides - useful for developing  
PT products for diagnosis, prevention and treatment of infections e.g.  
PT pneumonia, bacteremia, meningitis or endocarditis  
PS Claim 1: Page 122-123; 181pp; English.  
CC The sequence is that of a Streptococcal polypeptide coding region.  
CC The polypeptide can potentially be used for the diagnosis and  
CC prevention of bacterial infections, especially SP infection.  
CC It may be used for the treatment of diseases such as otitis media,  
CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural  
CC empyema, endocarditis or infection of the cerebrospinal fluid.  
CC Sequence 2130 BP; 639 A; 385 C; 498 G; 608 T;

Query Match 8.8%; Score 16; DB 1; Length 2130;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AACACAGCAAAATCA 175  
|  
Db 1369 AACACAGCAAAATCA 1354

RESULT 5  
T98762  
ID T98762 standard; DNA; 529 BP.  
AC T98762;  
DT 10-NOV-1998 (first entry)  
DE DNA encoding a S. pneumoniae PFS protein.  
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
KW immunological response; inoculation; antibody production; inhibitor;  
KW T cell immune response; antimicrobial compound; bacterial adhesion;  
KW extracellular matrix protein; protein-mediated cell invasion; wound;  
KW pathogenesis; ss.  
OS Streptococcus pneumoniae.  
FH Key Location/Qualifiers  
FT CDS Complement(40..234)  
FT /tag- a  
PN MO9743303-A1.  
PD 20-NOV-1997.  
PF 14-MAY-1997; U07950.  
PR 14-MAY-1996; US-017670.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,  
PI Stedola RK:  
DR WPI: 98-008793/01.  
DR P-PSDS: w38720.  
PT Novel Streptococcus pneumoniae proteins and related DNA - useful for  
PT diagnosing anti-microbial agents for treatment of bacterial  
PT infections  
PS Claim 4; Page 244; 483pp; English.  
CC This sequence encodes a Streptococcus pneumoniae protein that (based on  
CC homology with an Escherichia coli protein) is a pfs protein, and  
CC represents a DNA sequence of the invention. The DNA sequences were  
CC isolated from Streptococcus pneumoniae strain 0100993 (NCIMB 40794). The  
CC Streptococcus pneumoniae proteins of the invention can be used to  
CC identify compounds which interact with and inhibit or activate the  
CC activity of the proteins. Antagonists can be used to treat diseases  
CC caused by S. pneumoniae proteins, through genetic immunisation. They can  
CC also be used to induce an immunological response in a mammal by  
CC inoculation with the S. pneumoniae proteins or delivery of the encoding  
CC nucleic acids in a vector adequate to produce antibody and/or T cell  
CC immune responses to protect the animal from disease. The proteins can  
CC also be used to identify antimicrobial compounds which are capable of  
CC inhibiting their bioactivity. In particular the proteins of the invention  
CC can be used to prevent adhesion of bacteria to mammalian extracellular  
CC matrix proteins on in-dwelling devices or in wounds, to block protein-  
CC mediated mammalian cell invasion, and to block the normal progression of  
CC pathogenesis in infections initiated other than by the implantation of  
CC in-dwelling devices or other surgical techniques.

SQ Sequence 529 BP; 173 A; 111 C; 89 G; 156 T;

Query Match 8.8%; Score 16; DB 1; Length 529;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AACACAGCAAAATCAA 175  
|||||  
DB 271 AACACAGCAAAATCAA 286

RESULT 6  
ID V59890/C  
AC V59890 standard; DNA: 829 BP.  
DT 24-NOV-1998 (first entry)  
DE DNA encoding Staphylococcus aureus xanthine permease proteins.  
KW Bacillus subtilis; xanthine permease; transport protein; treatment;  
KW prevention; bacterial infection; Helicobacter pylori; vaccine; ss.  
OS Staphylococcus aureus.  
FH Key Location/Qualifiers  
FT CDS complement (1..636)  
FT FT /\*tag= a  
FT FT /note= "no termination codon given; encodes W79363"  
FT CDS complement (637..795)  
FT FT /\*tag= b  
FT FT /note= "no termination codon given; encodes W79364"

M09823738-A2.  
PN 04-JUN-1998.  
PE 24-NOV-1997; U22092.  
PR 25-NOV-1996; US-031469.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Warren RL:  
DR WPI: 98-322718/28.  
DR P-PSDS: W79363; W79364.  
PT New nucleic acid from Staphylococcus aureus NCIMB 40771 - useful  
PT for, e.g. diagnosis; prevention and treatment of bacterial  
PT infection(s)  
PS Claim 4; Page 60; 114pp; English.  
CC The present sequence encodes Staphylococcus aureus WCHU (NCIMB 40771)  
CC proteins that have homology to a Bacillus subtilis xanthine  
CC permease and transport protein. The proteins are used to generate  
CC antibodies and to screen for antimicrobials. The products are used  
CC to treat or prevent bacterial infections, particularly

where caused by *S. aureus* but also against *Helicobacter pylori*.  
CC Particular applications are to treat subjects before surgery or  
CC insertion of an indwelling device (alternatively the device itself is  
CC impregnated before placement). The nucleic acid sequence is used as  
CC sources of antisense sequences (for therapeutic use) or regulatory  
CC elements for controlling expression of bacterial genes, and for  
CC antibacterial screening. The protein can be also used as a vaccine.  
SO Sequence 829 BP; 263 A; 177 C; 126 G; 243 T;

Query Match 8.8%; Score 16; DB 1; Length 829;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 16 CAATAAGTAAACAGCA 31  
|||||  
Db 634 CAATAAGTAAACAGCA 619

RESULT 7  
V75326  
V75326 standard; DNA; 462 BP.  
ID V75326;  
AC 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #1015.  
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
OS Staphylococcus aureus.  
PN EP-786519-42.  
PD 30-JUL-1997.  
PF 07-JAN-1997; 100117.  
PR 05-JAN-1996; US-009861.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunesh CA,  
PI Rosen CA:  
DR WPI; 97/374922/35.  
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
PS Claim 1; Page 1794; 3271pp; English.  
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the S.aureus DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against S.aureus infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.  
SO Sequence 462 BP; 206 A; 75 C; 77 G; 104 T;

Query Match 8.8%; Score 16; DB 1; Length 462;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 161 AACAGCAAAATCAAA 176  
|||||  
Db 174 AACAGCAAAATCAAA 189

RESULT 8

V74617/C  
 ID V74617 standard; DNA: 6076 BP.  
 AC V74617;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #306.  
 KM Computer readable medium; vaccine; S. aureus infection; immunodetection;  
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KM skin infection; surgical wound infection; scalded skin syndrome;  
 KM toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 FH Key Location/Qualifiers  
 FT misc\_feature  
 FT 241..300  
 FT /tag= a  
 FT /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 FT misc\_feature  
 FT 2041..2100  
 FT /tag= b  
 FT /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 FT misc\_feature  
 FT 3841..3900  
 FT /tag= c  
 FT /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 FT misc\_feature  
 FT 5641..5700  
 FT /tag= d  
 FT /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"  
 PN EP-786519-A2.  
 PN 30-JUL-1997.  
 PF 07-JAN-1997: 100117.  
 PR 05-JAN-1996: US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA.  
 PI Rosen CA; 97-374922/35.  
 DR WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
 PT stored on computer readable medium and used in the production of  
 PT anti-S. aureus vaccines  
 PS Claim 1: Page 1142-1146: 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S. aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S. aureus infection. The S. aureus polypeptides can also be used in a kit for the immunodetection of S. aureus in a sample. S. aureus is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the S. aureus DNA sequences contained on the computer readable medium.  
 CC Sequence 6076 BP: 1861 A: 1175 C: 886 G: 1914 T:  
 SO

Db 4941 CAATTAAGTAAACAGGA 4926  
 ||||||||||||||||  
 RESULT 9  
 X30886/C  
 ID X30886 standard; DNA: 528 BP.  
 AC X30886;  
 DT 20-MAY-1999 (first entry)  
 DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:163.  
 KM Streptococcus pneumoniae strain 0100993; vaccine; immune response;  
 KM streptococcal infection; pneumococcal; ss.  
 OS Streptococcus pneumoniae.  
 FH Key Location/Qualifiers  
 FT STREPTOCOCCUS PNEUMONIAE  
 FT W09737/026-A1.  
 FT PD 09-OCT-1997.  
 FT PF 01-APR-1997: U05306.  
 FT PR 22-AUG-1996: US-025788.  
 FT PR 02-APR-1996: US-014690.  
 FT PA (SMK ) SMITHKLINE BEECHAM CORP.  
 FT PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,  
 FT PI Stodola RK;  
 FT DR WPI: 97-503111/46.  
 FT P-PSDB: Y11304.  
 PR Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
 PT vaccines, drug screening, etc  
 PS Claim 5: Page 153: 354pp; English.  
 CC X30724 to X30946 represent genomic DNA sequences isolated from  
 CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
 CC encode the novel proteins given in Y1114 to Y1167. The proteins,  
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
 CC streptococcal infections and in assays for identifying compounds that  
 CC inhibit or activate the activity of the proteins. The antagonists can  
 CC be used to treat an individual having need to inhibit a bacterial  
 CC protein. Vectors expressing the proteins can be used to induce a  
 CC protective immune response in mammals.  
 CC Sequence 528 BP: 158 A: 88 C: 109 G: 173 T:  
 SO

Query Match 8.8%; Score 16; DB 1: Length 528;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 160 AAACAAGCAAAATCAA 175  
 ||||||||||||||||  
 Db 255 AAACAAGCAAAATCAA 240  
 RESULT 10  
 X006001  
 ID 006001 standard; DNA: 12284 BP.  
 AC 006001;  
 DT 16-JAN-1991 (first entry)  
 DE Sequence encoding protein characteristic of hog cholera virus (HCV).  
 KM Vaccine; parvovirus; swine influenza virus.  
 OS Hog cholera virus.  
 FH Key Location/Qualifiers  
 FT HOG CHOLERA VIRUS  
 FT CDS 364..12057  
 FT /tag= a  
 FT EP-389034-A.  
 FT PD 26-SEP-1990.  
 FT PF 12-MAR-1990: 200573.  
 FT PR 19-MAR-1989: EP-104921.  
 FT PR 12-MAR-1990: EP-200573.  
 FT PA (ALKU ) AKZO NV.  
 FT PI Meyers G, Kumenaf T, Thiel HT;  
 FT DR WPI: 90-291979/39.  
 DR P-PSDB: R06996.  
 PT New hog cholera virus vaccine and diagnostic - comprises nucleic acid sequence of poly-peptide characteristic of hog cholera virus  
 PS Claim 2: Fig 2: 28pp; English.  
 CC Gene product may be used to provide a vaccine and Abs for diagnosis of hog cholera viral infection in pigs.  
 CC

Sequence 12284 BP; 3793 A; 2601 C; 3228 G; 2662 T;

Query Match 8.3%; Score 15; DB 1; Length 12284;

Best Local Similarity 100.0%; Pred. No. 68;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 132 CAGAACAAACCACA 146

DB 9991 CAGAACAAACCACA 10005

RESULT 11

ID Q10350/c

AC Q10350; 8.3%; Score 15; DB 1; Length 12284;

DE 15-APR-1991 (first entry)

DE PLUXSPAFc-1 encoding Igg-binding/luciferase fusion protein.

KW Igg; Fc region binding domain; Protein A; luciferase; PLUXSPAFc-1;

FT ss. Location/Qualifiers

FT Key 16.1275

FT cds /\*tag- a

FT 1302..2276

FT /\*tag- b

PN J02308791-A.

PD 21-DEC-1990.

PE 24-MAY-1989; 131195.

PR 24-MAY-1989; JP-131195.

PA (SEKI ) SEKISUI CHEM IND KK.

DR WPI: 91-041061/06.

PT Polypeptide for fused protein prodn. - has binding ability to

PT Immunoglobulin Fc region and luciferase activity

PS Claim 2; Fig 7; 18pp; Japanese.

CC DNA encoding Protein A and having BamHI sticky ends is ligated to

CC BamHI-digested plasmid pDR720 to give pSPAFc-1. A sequence

CC encoding luciferase was isolated from V.harveyi DNA and ligated to

CC BamHI-digested pUC18 to give pLUX1801. A series of plasmids was

CC derived from this plasmid, including pLUX1803 which was digested

CC with HindIII and ligated with HindIII-digested pSPAFc-1. Plasmids

CC PLUXSPAFc-1 and -2 were obtained. They encode fusion proteins for

CC use as a second labelled antibody. See also Q10349 and Q10351.

SO Sequence 2642 BP; 761 A; 559 C; 627 G; 695 T;

Query Match 8.3%; Score 15; DB 1; Length 2642;

Best Local Similarity 100.0%; Pred. No. 66;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 CAAACCAACCAACAA 151

DB 1189 CAAACCAACCAACAA 1175

RESULT 12

ID Q10351/c

AC Q10351; 8.3%; Score 15; DB 1; Length 1153;

DE 15-APR-1991 (first entry)

DE PLUXSPAFc-2 encoding Igg-binding/luciferase fusion protein.

KW Igg; Fc region binding domain; Protein A; luciferase; PLUXSPAFc-2;

FT ss. Location/Qualifiers

FT Key 16.1266

FT cds /\*tag- a

FT 1293..2267

FT /\*tag- b

PN J02308791-A.

PD 21-DEC-1990.

PE 24-MAY-1989; 131195.

PR 24-MAY-1989; JP-131195.

PA (SEKI ) SEKISUI CHEM IND KK.

DR WPI: 91-041061/06.

PT Polypeptide for fused protein prodn. - has binding ability to

PT Immunoglobulin Fc region and luciferase activity

PS Claim 3; Fig 8; 18pp; Japanese.

CC DNA encoding Protein A and having BamHI sticky ends is ligated to

CC BamHI-digested plasmid pDR720 to give pSPAFc-1. A sequence

CC encoding luciferase was isolated from V.harveyi DNA and ligated to

CC BamHI-digested pUC18 to give pLUX1801. A series of plasmids was

CC derived from this plasmid, including pLUX1803 which was digested

CC with HindIII and ligated with HindIII-digested pSPAFc-1. Plasmids

CC PLUXSPAFc-1 and -2 were obtained. They encode fusion proteins for

CC use as a second labelled antibody. See also Q10349 and Q10350.

SO Sequence 2633 BP; 755 A; 561 C; 627 G; 690 T;

Query Match 8.3%; Score 15; DB 1; Length 2633;

Best Local Similarity 100.0%; Pred. No. 66;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 CAAACCAACCAACAA 151

DB 1180 CAAACCAACCAACAA 1166

RESULT 13

ID Q11661

AC Q11661; 8.3%; Score 15; DB 1; Length 1153 BP.

DE 09-JUL-1991 (first entry)

DE Complete human esterase D encoding CDNA.

KW Human esterase D; retinoblastoma; Wilson's disease; chromosome 13;

OS Homo sapiens.

FT Key Location/Qualifiers

FT cds 130..978

FT /\*tag- a

FT /product- human esterase D

PN US501773-A.

PD 30-APR-1991.

PE 31-AUG-1987; 091547.

PR 31-AUG-1987; US-091547.

PA (REGC ) UNIV OF CALIFORNIA.

PI Lee W-H, Lee Y-HP;

DR WPI: 91-140227/19.

DR P-FSDB: R11871.

PT cDNA sequence encoding human esterase D - used in prognosis for

PT genetic screening of retinoblastoma or Wilson's disease

PS Disclosure; fig 7; 29pp; English.

CC This is the sequence of a cDNA clone encoding complete recombinant

CC human esterase D protein. It can be used as a genetic marker and

CC hereditary or acquired genetic diseases associated with chromosome

CC 13 (locus 13q14.11). It can be used to determine individuals

CC predisposed to such diseases and for popn. screening. See also

CC Q11662, and Q11681.

SO Sequence 1153 BP; 361 A; 243 C; 231 G; 318 T;

Query Match 8.3%; Score 15; DB 1; Length 1153;

Best Local Similarity 100.0%; Pred. No. 66;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 ATTCTATAGTTGAAT 125

DB 1097 ATTCTATAGTTGAAT 1111

RESULT 14

ID Q11662

AC Q11662; 8.3%; Score 15; DB 1; Length 1069 BP.

DE 09-JUL-1991 (first entry)

DE Complete human esterase D encoding EL22 clone.

KW Human esterase D; retinoblastoma; Wilson's disease; chromosome 13;

KM ss. Location/Qualifiers  
 OS Homo sapiens.  
 FH Key 1..894  
 FT cds /\*tag= a  
 /product= human esterase D  
 PN US5011773-A.  
 PD 30-APR-1991.  
 PE 31-AUG-1987: US-091547.  
 PR 31-AUG-1987: US-091547.  
 PA (REGC ) UNIV OF CALIFORNIA.  
 PI Lee W-H, Lee T-HP;  
 DR WPI: 91-140227/19.  
 DR P-PSDB: R11872.  
 PT cDNA sequence encoding human esterase D - used in prognosis for  
 PT genetic screening of retinoblastoma or Wilson's disease  
 PS Disclosure; fig 8: 29pp; English.  
 CC This is the sequence of a cDNA clone, EL22, encoding recombinant  
 CC human esterase D protein. It can be used as a genetic marker and  
 CC diagnostic tool for retinoblastoma, Wilson's disease and other  
 CC hereditary or acquired genetic diseases associated with chromosome  
 CC 13 (locus 13q 14.11). It can be used to determine individuals  
 CC predisposed to such diseases and for popn. screening. See also  
 CC Q11661, and Q11681.  
 SQ Sequence 1069 BP: 348 A; 211 C; 211 G; 299 T;

Query Match 8.3%; Score 15; DB 1; Length 1069;

Best Local Similarity 100.0%; Pred. No. 66;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ATTCTATAGTGAAT 125  
 |||  
 Db 1013 ATTCTATAGTGAAT 1027

## RESULT 15

Q43504/C  
 ID 043504 standard; DNA; 1498 BP.  
 AC 043504;  
 DT 01-OCT-1993 (first entry)  
 DE Sequence encoding haemolysin protein.  
 KM Swine dysentery; vaccine; haemolysin protein; ss.  
 OS Treponema hyodysenteriae, strain B 204.  
 FH Key Location/Qualifiers  
 FT cds 1..456  
 FT /\*tag= a  
 FT /note= "unknown protein"  
 FT 471..1142  
 FT /\*tag= b  
 FT /product= haemolysin  
 FT cds  
 AU9227224-A.  
 PD 29-APR-1993.  
 PE 22-OCT-1992: 027224.  
 PR 25-OCT-1991: EP-202766.  
 PR 24-JUL-1992: EP-202774.  
 PA (DUIN ) DUPHAR INT RES BV.  
 PI Koopman MBH, Kusters JG, Mulr SJ, Koopmans M;  
 DR WPI: 93-188946/24.  
 DR P-PSDB: R38064, R40751.  
 PT polynucleotide encoding the haemolysin protein of Treponema  
 PT hyodysenteriae - used for obtaining polypeptide(s) for use in  
 PT vaccines for combating swine dysentery  
 PS Claim 2; Page 24-26; 41pp; English.  
 CC DNA encoding TH polypeptides was isolated by cloning TH DNA and  
 CC screening for haemolytic colonies. Q43504 encodes haemolysin (R40751)  
 CC and an unknown protein (R38064). It is contained in cell line E.  
 CC coli JM105(pJBA) [CBS 512.91]. Unlike the hemolysins from E. coli  
 CC recombinant T. hyodysenteriae hemolysin is calcium independent.  
 CC Plasmid pJBA containing the hemolysin gene tly of T. hyodysenteriae  
 CC was used in the PCR to obtain a DNA fragment containing the entire  
 CC reading frame of the gene of interest. Primers based on nucleotides  
 CC 471 to 490 and 1449 to 1430 were used to amplify a fragment

CC containing the tly gene sequence (see Q43505, Q43506).  
 SQ Sequence 1498 BP: 559 A; 171 C; 218 G; 550 T;

Query Match 8.3%; Score 15; DB 1; Length 1498;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 AGAAGGCTTCTATAA 92  
 |||  
 Db 1256 AGAAGGCTTCTATAA 1242

Search completed: August 5, 1999, 17:39:52  
 Job time: 6227 sec



Mon Aug 9 13:27:06 1999

us-09-049-696-4\_1.rng

---

Page 7

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:33:30 ; Search time 3268.17 Seconds  
(without alignments)  
176.134 Million cell updates/sec

Title: US-09-049-696-4

Sequence: 1 CAAAAGATGCACATTCATA.....ACACCAAAATCAAAATGC 181

Scoring table: OLIGO\_WUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: \*  
1: gb\_dal: \*  
2: gb\_dal: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_p11: \*  
8: gb\_p12: \*  
9: gb\_p11: \*  
10: gb\_p12: \*  
11: gb\_p13: \*  
12: gb\_p13: \*  
13: gb\_p13: \*  
14: gb\_p13: \*  
15: gb\_p13: \*  
16: gb\_p13: \*  
17: gb\_p13: \*  
18: gb\_p13: \*  
19: gb\_p13: \*  
20: gb\_p13: \*  
21: gb\_p13: \*  
22: gb\_p13: \*  
23: gb\_p13: \*  
24: gb\_p13: \*  
25: gb\_p13: \*  
26: gb\_p13: \*  
27: gb\_p13: \*  
28: gb\_p13: \*  
29: gb\_p13: \*  
30: gb\_p13: \*  
31: gb\_p13: \*  
32: gb\_p13: \*  
33: gb\_p13: \*  
34: gb\_p13: \*  
35: gb\_p13: \*  
36: gb\_p13: \*  
37: gb\_p13: \*  
38: gb\_p13: \*  
39: gb\_p13: \*  
40: gb\_p13: \*  
41: gb\_p13: \*  
42: gb\_p13: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result 8  
No. Score Match Length DB ID Description  
1 156 86.2 3311 11 AF039400 Homo sapi AF039400 Homo sapi

## SUMMARIES

2	156	86.2	2826	42	AF127036	Homo sapi
3	90	45.7	35278	11	AF039401	Homo sapi
4	20	11.0	126732	42	AC007114	Homo sapi
5	19	10.5	36616	36	CEFA6B6	270780 Caenorhabdi
6	19	10.5	1616	36	EOY17505	Y17505 Euplotes oc
7	19	10.5	1765	36	EOY18433	Y18433 Euplotes oc
8	18	9.9	1424	3	SSMAPROC	X52566 Pig mRNA fo
9	18	9.9	114418	8	ATAC007047	AC007047 Arabidops
10	18	9.9	78154	11	AC002466	AC002466 Human BAC
11	18	9.9	103566	11	AC003087	AC003087 Human BAC
12	18	9.9	155382	11	AC004911	AC004911 Homo sapi
13	18	9.9	40247	11	HSCOS10	AF109718 Homo sapi
14	18	9.9	2937	12	AB017156	AB017156 Mus muscu
15	18	9.9	201175	34	AC004676	AC004676 Homo sapi
16	18	9.9	158262	35	AC007739	AC007739 Homo sapi
17	17	9.4	253505	1	AP000007	AP000007 Pyrococcu
18	17	9.4	86065	7	AB005243	AB005243 Arabidops
19	17	9.4	71522	7	AB006703	AB006703 Arabidops
20	17	9.4	95266	7	ATP18E5	AL022603 Arabidops
21	17	9.4	88164	7	ATP9A14	AL035636 Arabidops
22	17	9.4	2128	7	MCUS1563	U51563 Mesembryant
23	17	9.4	74494	8	ATAC007171	AC007171 Arabidops
24	17	9.4	4309	9	AB018285	AB018285 Homo sapi
25	17	9.4	293	9	HS48D6E	Z61228 H. sapiens C
26	17	9.4	97847	9	HS874C20	AL021997 Homo sapi
27	17	9.4	126807	10	HS391022	AL031377 Human DNA
28	17	9.4	106793	11	AC002538	AC002538 Homo sapi
29	17	9.4	163027	11	AC002549	AC002549 Homo sapi
30	17	9.4	175120	11	AC004687	AC004687 Homo sapi
31	17	9.4	126041	11	AC004904	AC004904 Homo sapi
32	17	9.4	118788	11	AC005277	AC005277 Homo sapi
33	17	9.4	187543	11	AC005678	AC005678 Homo sapi
34	17	9.4	135521	11	AC005858	AC005858 Homo sapi
35	17	9.4	88036	11	HUAC003119	AC003119 Human Chr
36	17	9.4	8756	12	AF135060	AF135060 Rattus no
37	17	9.4	2404	12	MMCD14	X13987 Mouse CD14
38	17	9.4	1873	12	MUSCD14A	M34510 Mouse monoc
39	17	9.4	1813	12	RNNKRA1B	X33234 Rat NKAA1 g
40	17	9.4	755	17	HTLVLTR1	Z31659 Human T-cell
41	17	9.4	757	17	STLVLTR	Y13347 Simian T-ce
42	17	9.4	110000	34	CEY102F5_2	Continuation (3 of
43	17	9.4	104377	34	CEY62H9	Z99005 Caenorhabdi
44	17	9.4	258862	34	HS606P22	AL034551 Homo sapi
45	17	9.4	124956	34	HSJ193N13	AL078600 Homo sapi

## ALIGNMENTS

RESULT 1  
AF039400 3311 bp mRNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hcc1c1) mRNA,  
DEFINITION complete cds.  
ACCESSION AF039400  
VERSION 94009457  
KEYWORDS AF039400.1 GI:4009457  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 3311)  
Gruber/A.D., Elble/R.C., Ji/H.L., Schreuer/R.D., Fuller/C.M. and  
Pauli/B.U.  
TITLE Genomic cloning, molecular characterization, and functional  
analysis of human hcc1c1, the first human member of the family of  
Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 3311)  
Gruber/A.D., Elble/R. and Pauli/B.U.  
AUTHORS Direct Submission  
TITLE

JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA

FEATURES  
source

1. .3311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
/tissue\_type="small intestine"  
1. .3311  
/gene="hCLCA1"  
/gene="hCLCA1"  
352. .3096  
/note="transmembrane glycoprotein"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95428.1"  
/db\_xref="GI:4009458"  
/translation="MGPKSSVFLLILHLLEGALNSLIQNNNGYEGIVVAIDPNVP  
EDTLIOIKDMVTQASLYLFKATGRKRYFNVAIILPETWKADYRKLLEYKNA  
DIVAESTPPNDPEYBOMNGCEGGERHILTPETGKLAEGPGKAFVEMAH  
LRMGVDEYNDEKFIYSNRIQAVRSAGITGNVYKKGCGSCYTRKCTENKVTGL  
YKGCERTYLSRQTERKASIMFAOHVDSIVFCTEONHKEAPKNOKNCMLASTWEYI  
RDSDEFKRTYPTOPNPTSLQIGQIVCLVLDKSGMAVGNRLNUNQAGLEL  
LQVEIKSPWGMVFDSAAHVOSELIOINSQSDRLAKRLPAASGGSTISCSLSRA  
FTVIRKRYPDGSEIYVLDGEDNTISGCENEGVSGAIHTVALGSPAAQLELSK  
MTGLOIYASDOVONNGILIDAFGALSNGAVSORSIOLESKGLTLOSOMNGTYIV  
DSYVGRNLTLLITWTPPOILLMPDSGOGGFVYDKNTKMAVLIQIGAKGTWKY  
SLQASQTLITVTSRASNATLPRTYTSKNTKSPPLVYANRGAAPILRA  
SVTALIESVNGKTYVLELDNGAGADTKDQVYSRTFTYDNGRISVYRALGYN  
AARVAVIPQOSGALYIPGMINDENIQNPPRINDVQHKVCFSTRSSGSEFVAS  
DVNAPRIPDLFPFGQIDLKAIEHIGSLINLTWAPGDYDHTAHKYLIRISTSID  
LRDKNESLQVNTTALIPKANESEVLFEPKENTIFENGTDLFAIQADVVDKLSI  
SNIAVSLFIPQTPPETPSPDETSAFCPNHINSTIPGILHILKIMMWIGELQLSIA

gene

CDS

BASE COUNT 1028 a 692 c 742 g 849 t  
ORIGIN

Query Match 86.2%; Score 156; DB 11; Length 3311;  
Best Local Similarity 100.0%; Pred. No. 1.5e-74;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ACAGGACTCTATGAAAAAGATGTGAGTTGTTCTCCAATCCCGCAGACGAGAAAGGCT 85  
DB 997 ACAGGACTCTATGAAAAAGATGTGAGTTGTTCTCCAATCCCGCAGACGAGAAAGGCT 1056  
QY 86 TCTATTAATGTTTGGCAACAATGTGATTTCTATAGTTGTAATCTGTACAGACAAACAC 145  
DB 1057 TCTATTAATGTTTGGCAACAATGTGATTTCTATAGTTGTAATCTGTACAGACAAACAC 1116  
QY 146 AACAAAGAAGCTCCAAACAGCAAAATCAAAAATGC 181  
DB 1117 AACAAAGAAGCTCCAAACAGCAAAATCAAAAATGC 1152

RESULT 2  
AF127036 2826 bp mRNA PRI 16-APR-1999  
LOCUS Homo sapiens calcium-activated chloride channel protein 1 (CACCL)  
DEFINITION mRNA, complete cds.  
ACCESSION AF127036  
NID 94585468  
VERSION AF127036.1 GI:4585468  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2826)  
AUTHORS Agnel, M., Vernet, T. and Culouscou, J.-M.  
TITLE Cloning of three human homologs of bovine epithelial chloride

JOURNAL channel  
REFERENCE 2 (bases 1 to 2826)  
AUTHORS Agnel, M. and Culouscou, J.-M.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des  
Carrières, Rueil-Malmaison 92500, France

FEATURES  
source

1. .2826  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="small intestine; colon"  
1. .2826  
/gene="CACCL1"  
5. .2749  
/gene="CACCL1"  
/note="bovine epithelial chloride channel homolog"  
/codon\_start=1  
/product="calcium-activated chloride channel protein 1"  
/protein\_id="AAD5467.1"  
/db\_xref="PID:94585469"  
/db\_xref="GI:4585469"  
/translation="MGPKSSVFLLILHLLEGALNSLIQNNNGYEGIVVAIDPNVP  
EDTLIOIKDMVTQASLYLFKATGRKRYFNVAIILPETWKADYRKLLEYKNA  
DIVAESTPPNDPEYBOMNGCEGGERHILTPETGKLAEGPGKAFVEMAH  
LRMGVDEYNDEKFIYSNRIQAVRSAGITGNVYKKGCGSCYTRKCTENKVTGL  
YKGCERTYLSRQTERKASIMFAOHVDSIVFCTEONHKEAPKNOKNCMLASTWEYI  
RDSDEFKRTYPTOPNPTSLQIGQIVCLVLDKSGMAVGNRLNUNQAGLEL  
LQVEIKSPWGMVFDSAAHVOSELIOINSQSDRLAKRLPAASGGSTISCSLSRA  
FTVIRKRYPDGSEIYVLDGEDNTISGCENEGVSGAIHTVALGSPAAQLELSK  
MTGLOIYASDOVONNGILIDAFGALSNGAVSORSIOLESKGLTLOSOMNGTYIV  
DSYVGRNLTLLITWTPPOILLMPDSGOGGFVYDKNTKMAVLIQIGAKGTWKY  
SLQASQTLITVTSRASNATLPRTYTSKNTKSPPLVYANRGAAPILRA  
SVTALIESVNGKTYVLELDNGAGADTKDQVYSRTFTYDNGRISVYRALGYN  
AARVAVIPQOSGALYIPGMINDENIQNPPRINDVQHKVCFSTRSSGSEFVAS  
DVNAPRIPDLFPFGQIDLKAIEHIGSLINLTWAPGDYDHTAHKYLIRISTSID  
LRDKNESLQVNTTALIPKANESEVLFEPKENTIFENGTDLFAIQADVVDKLSI  
SNIAVSLFIPQTPPETPSPDETSAFCPNHINSTIPGILHILKIMMWIGELQLSIA

gene

CDS

BASE COUNT 875 a 623 c 632 g 696 t  
ORIGIN

Query Match 86.2%; Score 156; DB 42; Length 2826;  
Best Local Similarity 100.0%; Pred. No. 1.5e-74;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 ACAGGACTCTATGAAAAAGATGTGAGTTGTTCTCCAATCCCGCAGACGAGAAAGGCT 85  
DB 650 ACAGGACTCTATGAAAAAGATGTGAGTTGTTCTCCAATCCCGCAGACGAGAAAGGCT 709  
QY 86 TCTATTAATGTTTGGCAACAATGTGATTTCTATAGTTGTAATCTGTACAGACAAACAC 145  
DB 710 TCTATTAATGTTTGGCAACAATGTGATTTCTATAGTTGTAATCTGTACAGACAAACAC 769  
QY 146 AACAAAGAAGCTCCAAACAGCAAAATCAAAAATGC 181  
DB 770 AACAAAGAAGCTCCAAACAGCAAAATCAAAAATGC 805

RESULT 3  
AF039401 35278 bp DNA PRI 15-DEC-1998  
LOCUS Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene,  
DEFINITION complete cds.  
ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Eible,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.O.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human hCICAI, the first human member of the family of Ca<sup>2+</sup>-activated Cl<sup>-</sup> channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Eible,R. and Pauli,B.O.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
Source 1. .35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1. .1617  
/gene="hCICAI"  
1. .33522  
/gene="hCICAI"  
767. .1265  
/gene="hCICAI"  
repeat\_region  
/rpt\_family="L1 transposable element"  
/rpt\_type="dispersed"  
1587. .1596  
/gene="hCICAI"  
/note="putative"  
1618. .1708  
/gene="hCICAI"  
/number=1  
join(1618. .1708,1962. .2383,6667. .6807,6979. .7126,9696. .9801,15454. .15631,18591. .18809,19783. .20010,22253. .22427,24522. .24628,26644. .26859,27447. .27708,28766. .28936,31840. .32079,32919. .33522)  
/gene="hCICAI"  
/product="calcium-dependent chloride channel-1"  
1962. .2383  
/gene="hCICAI"  
/number=2  
join(2222. .2383,6667. .6807,6979. .7126,9696. .9801,15454. .15631,18591. .18809,19783. .20010,22253. .22427,24522. .24628,26644. .26859,27447. .27708,28766. .28936,31840. .32079,32919. .33310)  
/gene="hCICAI"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="GI:4009460"  
/db\_xref="PID:94009460"  
/translation="MGPFKSSVFILILHLELEGALNSLIOLNNGYEGIVVAIDPNVP  
EDTLLIQIKDMYTOASLYLEFATGRFRYKNVAIILPEWTKRADYVRKLETTYKA  
DIVVAESTPGNDPEYTEOMGNGEGERIHLPDFIAGKLAIEYGVQAFYEHNA  
LRGVAFEDVFNNDKEFTLSNGRIQAVRCSAGITGTNYKKGGSCYTRKCTEMVGL  
YKGCETVLOSROTEKASIMFAQVDSIVFCFEONHKNKAPKQNKCLBSTWEYI  
RQSEDEKRTTPMTQPPNPTFSLQIGQRTVCIVLDKSGSMATGNRLNRLNQGLFL  
LQVELGSWGWMTFDSAANVOSELIOINGSDVRSKQSLIHTVALPSPAQOELELSK  
FTVIRKKYPTDGESEIYLIDGEDNTISGCNEVKSQALIHHTVALPSPAQOELELSK  
MTGGLOTVASDOVONNGLLIDAFGLASGNNAVRSQIESKGLTIONSQWNGTIV  
DSTVGRDTEFLITWTQPOIILMDPSGCGFVDRKSPPLVYANIRGASPTILRA  
SVIALIESVNGKVTLELLDNGAGADATKQDGYISRTFTYDNGRYSVNRVRLGGVN  
AARARVYPOOSGALYIPGWLENDEIQMNPBPINMDVQHKVCFSRSGSPVAS  
DAVNADIPDLFPPGOITDLNAELHGGSLMLTWTAPGDDYDHQTAHKYIIRISTSID  
LRKFMSLOVNTALIPKANSSEVFLFKPEVITFENGDTLITAIQAVDKVLSKI  
SNARVSLFIPQTPPETPSPDSETSAPCPHINSTIPGIHIKIMKMGELQLSIA"

exon 6667. .6807  
/gene="hCICAI"  
/number=3  
6979. .7126  
/gene="hCICAI"  
exon

exon /number=4  
9696. .9801  
/gene="hCICAI"  
exon /number=5  
15454. .15631  
/gene="hCICAI"  
exon /number=6  
18591. .18809  
/gene="hCICAI"  
exon /number=7  
19783. .20010  
/gene="hCICAI"  
exon /number=8  
22253. .22427  
/gene="hCICAI"  
exon /number=9  
24522. .24628  
/gene="hCICAI"  
exon /number=10  
26644. .26859  
/gene="hCICAI"  
exon /number=11  
27447. .27708  
/gene="hCICAI"  
exon /number=12  
28766. .28936  
/gene="hCICAI"  
exon /number=13  
31840. .32079  
/gene="hCICAI"  
exon /number=14  
32919. .33522  
/gene="hCICAI"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 49.7%; Score 90; DB 11; Length 35278;  
Best local Similarity 100.0%; Pred. No. 9.3e-39;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 26 ACAGAGCTCATGAAAGAGATGAGTTGTTCTCCATCCGCCGACGAGGAGAGCT 85  
DB 15542 ACAGAGCTCATGAAAGAGATGAGTTGTTCTCCATCCGCCGACGAGGAGAGCT 15601  
OY 86 TCTATATGTTTGCACAACATGTTGATTC 115  
DB 15602 TCTATATGTTTGCACAACATGTTGATTC 15631  
RESULT 4  
AC007114  
LOCUS AC007114 126732 bp DNA PRI 12-APR-1999  
DEFINITION Homo sapiens chromosome 17, clone hRP.166\_P.13, complete sequence.  
ACCESSION AC007114  
NID 94581193  
VERSION AC007114.7 GI:4581193  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 126732)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckert,J., Berr,M., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearlano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,

```

Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Hearford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoccky,J., Lieu,C., Locke,K., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (18-MAR-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 126732)
TITLE
JOURNAL
REFERENCE
AUTHORS
Bairren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barne,N., Beckelky,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Hearford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoccky,J., Lieu,C., Locke,K., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (12-APR-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 12, 1999 this sequence version replaced gi:4580489.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only 126732 base pairs from the middle of this clone are being
submitted. The remainder overlaps either accession number AC003950
(WICR project L261) or accession number AC004584 (WICR project
L312).
FEATURES
source
Location/Qualifiers
1. 126732
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRP.166_P.13"
/clone_lib="RPCI-11 human BAC library"
/map="17"
/chromosome="17"
complement(225..399)
/rpt_family="MIR"
634..635
/note="Unresolved compression. Probably CC."
repeat_region
1159..1195
/rpt_family="AT_rich"
repeat_region
3460..3489
complement(3490..3642)
/rpt_family="AT_rich"
repeat_region
4211..4233
complement(4234..4256)
/rpt_family="MIR104"
repeat_region
4243..4402
complement(4403..4425)
/rpt_family="(CA)n"
repeat_region
7083..7483
complement(7484..7506)
/rpt_family="AluDo"
repeat_region
7638..7882
complement(7883..7905)
/rpt_family="MER47B"
repeat_region
8018..8113
complement(8114..8136)
/rpt_family="AluSq"
repeat_region
8424..8563
complement(8564..8586)
/rpt_family="MIR5B"
/rpt_family="MIR5A"

Repeat_region complement(863..9093)
Repeat_region /rpt_family="AluSq"
Repeat_region 9133..9428
Repeat_region /rpt_family="AluYb8"
Repeat_region 9686..9790
Repeat_region /rpt_family="FLAM_A"
Repeat_region complement(10090..10230)
Repeat_region /rpt_family="LIM4"
Repeat_region 10237..10403
Repeat_region /rpt_family="MER5B"
Repeat_region 10410..10677
Repeat_region /rpt_family="AluXs"
Repeat_region 10809..10849
Repeat_region /rpt_family="MER57B"
Repeat_region complement(10858..11072)
Repeat_region /rpt_family="AluSp"
Repeat_region 11073..11093
Repeat_region /rpt_family="AT_rich"
Repeat_region complement(11257..11436)
Repeat_region /rpt_family="AluXs"
Repeat_region 11437..11475
Repeat_region /rpt_family="(TG)n"
Repeat_region complement(11476..11600)
Repeat_region /rpt_family="AluXs"
Repeat_region 12840..13143
Repeat_region /rpt_family="AluSq"
Repeat_region 14926..15020
Repeat_region /rpt_family="Charlies"
Repeat_region 16416..16439
Repeat_region /rpt_family="(A)n"
Repeat_region 16455..16733
Repeat_region /rpt_family="AluYb"
Repeat_region 16995..17296
Repeat_region /rpt_family="AluDo"
Repeat_region 17521..17655
Repeat_region /rpt_family="HAL1"
Repeat_region 17794..18087
Repeat_region /rpt_family="HAL1"
Repeat_region 18667..18830
Repeat_region /rpt_family="MER105"
Repeat_region complement(18868..19055)
Repeat_region /rpt_family="L2"
Repeat_region 19912..20197
Repeat_region /rpt_family="AluDo"
Repeat_region 20198..20224
Repeat_region /rpt_family="(TAA)n"
Repeat_region 21204..21281
Repeat_region /rpt_family="MIR"
Repeat_region 22304..22496
Repeat_region /rpt_family="MER3"
Repeat_region 22544..22564
Repeat_region /rpt_family="AT_rich"
Repeat_region 26544..26600
Repeat_region /rpt_family="GC_rich"
Repeat_region complement(26694..26995)
Repeat_region /note="Unresolved compression. Probably GC."
Repeat_region complement(27657..27959)
Repeat_region /rpt_family="AluY"
Repeat_region complement(27976..28200)
Repeat_region /rpt_family="L2"
Repeat_region 28331..28452
Repeat_region complement(28453..28752)
Repeat_region /rpt_family="AluXs"
Repeat_region 28753..28774
Repeat_region /rpt_family="L2"
Repeat_region 28775..28812
Repeat_region /rpt_family="(CA)n"
Repeat_region 28812..28832
Repeat_region /rpt_family="(TC)n"
Repeat_region 28813..28963
Repeat_region complement(28964..29487)

```

```
repeat_region /rpt_family="L2"
29488..29798 /rpt_family="AluSg"
repeat_region complement(29799..29961)
/rpt_family="L2"
repeat_region 30630..30657
/rpt_family="(TAAAAA)n"
repeat_region 31112..31405
/rpt_family="AluJo"
repeat_region 31424..31572
/rpt_family="AluJo"
repeat_region 31573..31874
/rpt_family="AluJo"
repeat_region 31875..32033
/rpt_family="AluJo"
repeat_region 32621..32794
/rpt_family="MLT1C"
repeat_region 32801..33062
/rpt_family="AluY"
repeat_region 33073..33368
/rpt_family="AluSx"
repeat_region 33397..33656
/rpt_family="MLT1C"
repeat_region complement(34645..34822)
/rpt_family="MER91A"
repeat_region 35047..35349
/rpt_family="AluJo"
repeat_region 35382..35777
/rpt_family="L1MD2"
repeat_region 35970..35990
/rpt_family="AT-rich"
repeat_region 36131..36311
/rpt_family="MIR"
repeat_region complement(36361..36434)
```

```
Query Match 11.0% Score 20; DB 42; Length 126732;
Best Local Similarity 100.0%; Pred. No. 0.87; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;
```

```
Y 135 AACAAACACACAAAGAA 154
|||
Db 9650 AACAAACACACAAAGAA 9669
```

```
RESULT 5
CEFA6B6/c 36616 bp DNA INV 04-DEC-1998
DEFINITION Caenorhabditis elegans cosmid F46B6, complete sequence.
ACCESSION 270780
VERSION 91263454
NID 270780.1 GI:1263454
KEYWORDS HTG: GFP-binding; Homeobox; Initiation factor IF-2; Lipase;
Transposon: Zinc finger.
SOURCE Caenorhabditis elegans.
```

```
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
Rhabditiina; Rhabditiidae; Rhabditiidae; Peloderiinae; Caenorhabditis.
1 (bases 1 to 36616)
```

```
REFERENCE 1 (bases 1 to 36616)
AUTHORS MCMURRAY,A.
TITLE Direct Submission
JOURNAL Submitted (16-APR-1996) Louis, MO 63110, USA. E-mail:
jess@sanger.ac.uk or twenematode.wustl.edu
2 (bases 1 to 36616)
```

```
REFERENCE 1 (bases 1 to 36616)
AUTHORS Wilson,R., Almscough,R., Anderson,K., Baynes,C., Berks,M.,
Burton,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulston,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kistler,J., Laister,N.,
Laitelle,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra,A.,
Saunders,D., Showkeen,R., Smalton,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterson,R., Watson,A., Weinstock,L.,
```

TITLE  
JOURNAL  
MEDLINE  
COMMENT

#### FEATURES

source

Wilkinson-Sproat,J. and Wohldman,P.  
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans  
Nature 368 (6466), 32-38 (1994)  
94150718  
Coding sequences below are predicted from computer analysis, using  
predictions from Genefinder (P. Green, U. Washington), and other  
available information.  
For a graphical representation of this sequence and its analysis  
see:  
http://webc.sanger.ac.uk/cgi-  
bin/display?db=wormbase&class=Sequence&object=F46B6  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we arrange for a small  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.  
This sequence is the entire insert of clone F46B6. The true right  
end of clone F58B6 is at 12346 in this sequence. The start of this  
sequence (1..104) overlaps with the end of sequence 270754.  
The end of this sequence (36513..36616) overlaps with the start of  
sequence 272305.

Location/Qualifiers

1..36616  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="v"  
/clone="F46B6"  
3459..5460  
/gene="F46B6.2"  
join(3459..3515,3749..3879,3928..4022,4072..4205,  
4824..4930,4974..5460)  
/gene="F46B6.2"  
/codon\_start=1  
/protein\_id="CAA94819.1"  
/db\_xref="PID:e1346818"  
/db\_xref="PID:g3877200"  
/db\_xref="GI:3877201"  
/db\_xref="SPTREMBL:Q20444"  
/translation="MASTGPKWSLTPPLAAVAFEPALGAGAHVACSLPFFLPFPI  
N SAGVLTAFHAFILRYPNRCDFYLQITCSISGFEFFSSIMETGICINERKVDYIKD  
GVCHGKRTIAMVSCNDLNLQISIDKSGPKEREMIRFETSLITLIGLH  
LICTLTFYSATETKIRLKYSHYOVVIFLIIIVSFHLRCCFFELPLPAIGHS  
LLOGIVSWRTGCHGSTTRALNIGAAFAVIMATRSFGICFAMFNNTVPHMTRECHW  
LSHREAYCRRVHFTHPTIDWLPQETERIAAIQVNLITSLSKDPNFCQIMDVPTC  
QF"  
complement(6023..8771)  
/gene="F46B6.3"  
complement(join(6023..6184,6233..6342,6417..6498,  
6547..6693,7692..8200,8253..8578,8632..8771))  
/gene="F46B6.3"  
/note="cDNA EST yk465d10.3 comes from this gene; cDNA EST  
yk465d10.5 comes from this gene; cDNA EST yk481d9.5 comes  
from this gene"  
/codon\_start=1  
/protein\_id="CAA94820.1"  
/db\_xref="PID:e1346819"  
/db\_xref="PID:g3877201"  
/db\_xref="GI:3877201"  
/db\_xref="SPTREMBL:Q20445"  
/translation="MTDSKDGAVKVVLRRLPKYTEHEVLEQISPLPEVIGTYFHPA  
NESPFCARATLIVNPESECDMSMEFRRFDGIVFVDSGNSAAVVEASNONAKC  
DNRRKEDRVAALITDKYIDDFCKLEBERAIPITLLEQRIKLNOPDPAQIDKM  
EPLVLRFEKETGRDIDARQRDERAKRKRVDFEAKVNLDLKPSVPA  
SISATSKKDLKREKPKTEREKERKQDARKENMKRQFLDEKKKKHEERQDC  
PVPAPDKKREKRPSPKPAKVEQGEEDWIKLTPKSPAPKKKHPPFLALDLKC  
QKGTDLAPDDTDSFHHMKLKCNCGAPPHQVYVLENLADVGSGEANTLVK  
KCGVNTLITVEDFAFKYNTIQNKQKQIAFPCDRLGLEPFPDPRDENIAKSVETGN  
AHEIDLSERKAVDDDKRAKEAVEISEMSQFTTTRDKPK"

```
gene complement(8914..10546)
CDS /gene="F46B6.4"
complement(join(8914..9240,9289..9423,9511..10222,
10269..10439,10485..10546))
/gene="F46B6.4"
/note="similar to GTP-binding protein; cDNA EST yk255b11.3
comes from this gene; cDNA EST yk255b11.5 comes from this
gene"
/codon_start=1
/protein_id="CAA94821.1"
/db_xref="PID:e1346820"
/db_xref="GI:3877202"
/db_xref="GI:3877202"
/db_xref="SPTREMBL:Q20446"
/translation="MEILRSQTVLSEPTNLSIGSEVAAPSFAFANDRSVLYVAPKY
RMGSSASVLYQADROLEAVLYNDLNMNVAIDLIPVDYNTFRKRAVMSAGNEKL
IARRAARATLAVNVDALSPQOQELIRIPEVYFEDKXNYLAFKQFAPTEARLQ
IAIAEIPYIKRILHALSSKRHSRPDIILHDSYDLDGDLNELLRKQEDRKELKD
VTRKNGOIGVANSDDAVAVVGYNSGKTSYKRLTGAASLTPKQDLFALDITRL
AKLPGRSAVFDTIGFLSDLPMLHAAFEALIAHKASADVIHLRDISNDPKAQEE
DVLATLKSIGVDYLNERTISVDKIKRESAFPTSESNVSRIKCTGDGHELDIV
INDKYMVTKCTIRLIDARSPVEMLYHNLVIEPTIDGNYLIFPVWMESEIGR
FRKFAHLKKNSQSVSL"
complement(10993..13979)
/gene="F46B6.5"
complement(join(10993..11187,11234..11445,11651..11738,
11783..12355,12453..12561,12607..12857,12914..13120,
13164..13874,13932..13979))
/gene="F46B6.5"
/note="cDNA EST CEMS46F comes from this gene; cDNA EST
EMBL:D32534 comes from this gene; cDNA EST EMBL:D33154
comes from this gene; cDNA EST EMBL:D35195 comes from this
gene; cDNA EST EMBL:D35911 comes from this gene; cDNA EST
EMBL:D65058 comes from this gene; cDNA EST EMBL:D65320
comes from this gene; cDNA EST EMBL:D68240 comes from this
gene; cDNA EST EMBL:D68354 comes from this gene; cDNA EST
EMBL:D68702 comes from this gene; cDNA EST EMBL:C10986
comes from this gene; cDNA EST EMBL:C10765 comes from this
gene; cDNA EST EMBL:C12594 comes from this gene; cDNA EST
yk501h9.3 comes from this gene; cDNA EST yk348d2.3 comes
from this gene; cDNA EST yk348d2.5 comes from this gene;
cDNA EST yk195h1.3 comes from this gene; cDNA EST
yk195h1.5 comes from this gene; cDNA EST yk320d7.5 comes
from this gene"
/codon_start=1
/protein_id="CAA94828.1"
/db_xref="PID:e1358313"
/db_xref="PID:g3979815"
/db_xref="GI:3979815"
/translation="KGLILLIYGYIRRMGNQADAMQESPFGESISRPDNNRP
IYVLDRLDHTLEDITITFNDRGFRVKSVERGEDRLRLNRTITITSPASSGR
STSHOOLYGRSAVSRHKOQOYSTIRAPALAHMELQRYNQNADAREEQFLRVN
VLSVAPATSPVIVHOQLQASPAIQKQNVDEPNOSQITSGDLHOQEDSOK
TSBRKINPNONLAKVENALNPVGINHEGADIDPEVTESELEINAEFLRALDY
PEREMQDOODOPENEFGEVLEINELNASTIIOVEVEVEVEEGEVEGEIEAS
NRTVSGVPEPLSPIPKDLSPRMQGTTPKRGVQVQPKADNAPKRAAATDPRK
LDRERILRRSGRSRSLRYGSSPQSRPVEQKRHDTEREKRDHEKTTDPRK
REKRDHESDVYSHTITRHRYSEAHSTVQOSRESISTSSKSNAPEREKDAE
KERKRDKEHRLALEADNNKRYAREKOKOPSFHOKSNEOQISEIPVIRITQVGS
DKPKDRKOPIGSPSASAKRHDDVAREKPIKPKHSSSKPDRSSSSHPESSAMQ
SKPYSSYTKSADDELCDGDRKSDTINLPYGLARLSLLVNDCTTYOOFHDLQ
EKTSKRRSANLDSLSPANSEYSHNNITVNNHPKQSFAPRKGLALKSPKSPSS
SHSTDESSSCASSDNETSDATSQRSVANSPPAKKTRLSLDSKAEKMLDTHGPK"
complement(14372..17796)
/gene="F46B6.6"
complement(join(14372..14594,14641..15047,15877..17010,
17089..17212,17786..17796))
/gene="F46B6.6"
/note="similar to initiation factor IF-2; cDNA EST
CEMSD25F comes from this gene"
/codon_start=1
/protein_id="CAA94822.1"
/db_xref="PID:e1346821"
/db_xref="GI:3877203"
/db_xref="SPTREMBL:Q20447"
/translation="MFARLEAALSDVLSITEFELESIDRRNYDAIISDPLDVSIL
OTTAASFPRPVNGPKRKTEKEDADLHQLLESPDLERPPVITIMGVHDGKTL
LDARNSQIAGEFGGTTGHTGAFSEVLTGRVYTLDTGTGAAPASMRAGKADI
VVLVADDSVQKQTAOSIFARDANQVLAVANKIDKPNADPMKARSLLEHDVYE
OLGDDVQCEVESALQSRNLSLDALLQADVLDLAKTKGKEAVIETSSVHGIG
VCTLYVSRGTLKGCYLVAGSSWC RVATMDENGKIVQVATPSQPVAVSGMKDLPK
GDVLEAEVDRACQAVNRIDNMNREKADMDKEDQDKARAYLISNRKQDLDPG
ORRGSTLRNIVHKRIEKVDAGSPFRLIRTDVGTLEALILEISTKSOCKLO
INDPEVPPTEKQIELAKEDATLYTGNVTSKIKORAAAGIIDSENVYIRMES
LVGELSAKRIPLRLBELHVGSHVKEVYISDRGRKQPIGTYLNNGNPFRHITYFT
RGQNVYGEVEETMKAGMEVVTAKTITFEVGLALSDNNIRFKEDDOVEAEKRTITIOE
VMTXPPGF"
complement(18544..20673)
/gene="F46B6.7"
complement(join(18544..18609,18654..18845,18898..19198,
19249..20125,20175..20261,20305..20506,20554..20673))
/gene="F46B6.7"
/note="similar to zinc finger, C2H2 type; cDNA EST
EMBL:D27516 comes from this gene; cDNA EST EMBL:D27515
comes from this gene; cDNA EST CEMS82F comes from this
gene; cDNA EST CEMS82R comes from this gene; cDNA EST
EMBL:T02163 comes from this gene; cDNA EST EMBL:D32511
comes from this gene; cDNA EST EMBL:D34747 comes from this
gene; cDNA EST EMBL:D34747 comes from this gene; cDNA EST
```

```
Query Match 10.5%; Score 19; DB 36; Length 3616;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TGAAGAAGATGTGAGTT 55
DB 30607 TGAAGAAGATGTGAGTT 30589

RESULT 6
LOCUS E0Y17505 1616 bp DNA INV 17-JUN-1998
DEFINITION Euplotes octocarinatus gene encoding pheromone 5.
ACCESSION Y17505
NID 93687407
VERSION Y17505.1 GI:3687407
KEYWORDS pheromone 5.
SOURCE Euplotes octocarinatus.
ORGANISM Euplotes octocarinatus.
REFERENCE 1 (bases 1 to 1616)
AUTHORS Moellenbeck,M. and Heckmann,K.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1616)
AUTHORS Moellenbeck,M.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-1998) M. Moellenbeck, Institut fuer Allgemeine
zoologie und, Genetik, Schlossplatz 5, D-48149 Muenster, FRG
FEATURES
source location/Qualifiers
1..1616
1.1616
/organism="Euplotes octocarinatus"
/macronuclear
/strain="DM-5"
/db_xref="taxon:5937"
<163..228
/number=1
join(163..228,600..882,931..978,1026..1111)
/codon_start=1
/transl_table=10
/codon=(seq:"tga",aa:Cys)
/product="pheromone 5"
/protein_id="CAA76771.1"
/db_xref="PID:e1289594"
/db_xref="PID:g3687408"
/db_xref="GI:3687408"
```



```

exon          exon      743..1028  
                  /number=2  
Intron        Intron    1029..1078  
                /number=2  
exon          Exon     1079..1126  
                /number=3  
Intron        Intron   1127..1171  
                /number=3  
exon          Exon     1172..1257  
                /number=4  
  
BASE COUNT       584 A         280 G         605 T  
ORIGIN  
  
Query Match               10.5% Score 19; DB 36; Length 1765;  
Best local similarity 100.0%; Pident No. 3.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0:  
  
QY      159 CAACAGCAAAATCAAATAA 177  
              |||||  
DB      1288 CAAACAAGCAAAAATCAAATA 1306  
  
RESULT      8  
LOCUS       SSHAPROC           1424 bp      mRNA            MAM             23-MAR-1995  
DEFINITION Pig mRNA for haptocorrin (partial).  
ACCESSION   X52566  
NID         Q1963  
VERSION     X52566.1 GI:1963  
KEYWORDS    haptocorrin.  
SOURCE      pig.  
ORGANISM    Sus scrofa  
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Artiodactyla; Suidformes; Suina; Sulidae; Sus.  
REFERENCE   1 (bases 1 to 1424)  
AUTHORS     Alpers,D.  
TITLE       Direct Submission  
JOURNAL     Submitted (05-APR-1990) Alpers D., Washington University Medical School,  
                        Box 8124, Dept. Of Medicine, 660 S.Eucclid Ave, St Louis MO  
                        63110, USA  
            2 (bases 1 to 1424)  
REFERENCE   Hewitt,J.E., Seetharam,B., Leykam,J. and Alpers,D.H.  
AUTHORS     Isolation and Characterization of a cDNA encoding porcine gastric  
JOURNAL     haptocorrin  
MEDLINE     Eur. J. Biochem. 189 (1), 125-130 (1990)  
COMMENT     Data kindly reviewed (02-AUG-1990) by Alpers D.  
FEATURES  
Source  
          1. .1424  
            /organism="Sus scrofa"  
            /db_xref="taxon:9823"  
            /tissue_type="stomach"  
            /cclone_lib="lambda gt11"  
          <. .72  
            /product=""signal peptide (AA -24 to -1)"  
          <1. .1251  
            /note=""precursor protein (partial) (AA -24 to 392)."       
              /codon_start=-1  
              /protein_id="CMA36800.1"  
              /db_xref="PID:g1964"  
              /db_xref="GI:1964"  
              /translatio="ROSHOLPLVGLLFSLLPSQLCOSCVSEKDYSHLRLLISAMDND  
LEDIRGIVGAAILSLSORLAGIOWNSPEELRSORDDMRDMSNLTSGCALALLIAF  
GAKTDPVRRIHRRHHLYEKGEKFKEIKNNETINSRLNYILSYLSDEVYLCLFRNRN  
YSISENVTHFNPERKNENNGSGHSVDIGAVALVLTCVNKSISNGKIKAATKSDDTIQ  
VFESLVPAQAIIIPALIGKTLDVTKLTLVEKVGVNIIDEVPVPIISPENISVIYCVC  
KIEISNCINIITVELDMPKAAOENSNSTIYGFTMETPWPGPITSVGCIWANNERTWTM  
EHEDOOGITPRSMGTMLSMESI"  
          73. .1248  
            /product=""mature haptocorrin (AA 1 to 392)"
```

misc\_feature 1356..1360 /note="ATTTA box associated with RNA degradation"  
misc\_feature 1413..1418 /note="polya signal"  
BASE COUNT 440 a 326 c 251 g 367 t  
ORIGIN  
Query Match 9.9%; Score 18; DB 3; Length 1424;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 TGCACATTCATAAGTA 25  
1404 TGCACATTCATAAGTA 1421  
RESULT 9  
LOCUS ATAC007047/c DNA PLN 30-MAR-1999  
DEFINITION Arabidopsis thaliana chromosome II BAC F16F14 genomic sequence,  
complete sequence.  
ACCESSION AC007047  
NID 94544381  
VERSION AC007047.5 GI:4544381  
KEYWORDS HTG.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eustoids II; Brassicales; Brassicaceae;  
Arabidopsis.  
1 (bases 1 to 114418)  
Lin.X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,  
Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronging,C.M., Benito,M.,  
Carreira,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,  
Fraser,C.M. and Venter,J.C.  
Arabidopsis thaliana chromosome II BAC F16F14 genomic sequence  
Unpublished  
2 (bases 1 to 114418)  
Lin.X. and Kaul,S.  
Direct Submission  
Submitted (09-MAR-1999) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA, xlinetlgr.org  
3 (bases 1 to 114418)  
Lin.X.  
Direct Submission  
Submitted (30-MAR-1999) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
On Mar 30, 1999 this sequence version replaced gi:4510336.  
Address all correspondence to:  
Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr.  
Rockville, MD 20850, USA  
e-mail: xlinetlgr.org  
BAC clone F16F14 is from Arabidopsis chromosome II and is near the  
molecular marker m1398.  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.  
Genes were identified by a combination of three methods: Gene  
Prediction programs including GRAIL (available by anonymous ftp  
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of  
Washington), Genscan (Chris Burge,  
http://genomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene  
(http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the  
complete sequence against a peptide database and the Arabidopsis  
EST database at TIGR (http://www.tigr.org/tdb/at.html).  
Annotated genes are named to indicate the level of evidence for  
their annotation. Genes with similarity to other proteins are named  
after the database hits. Genes without significant peptide  
similarity but with EST similarity are named as 'unknown' proteins.  
Genes without protein or EST similarity, that are predicted by more

than two gene prediction programs over most of their length are  
annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
predicted by tRNAscan-SE (Sean Eddy,  
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are  
identified by RepeatMasker (Arian Smit,  
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of  
genomic sequence that are not annotated as genes but have predicted  
exons by GRAIL are annotated as misc features.  
FEATURES  
source  
Location/Qualifiers  
1..114418  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="II"  
/map="m1398"  
/clone="F16F14"  
join(<1..22,503..>2420)  
/gene="F16F14.1"  
<1..>2420  
/gene="F16F14.1"  
/note="predicted by genscan and genefinder"  
join(1..22,503..2420)  
/gene="F16F14.1"  
/note="hypothetical protein, 5' partial"  
/codon\_start=3  
/protein\_id="AAD22293.1"  
/db\_xref="PID:94544382"  
/db\_xref="GI:4544382"  
/translation="IQRMRIFLMTSVDEDEDEDETPFSLMDSERASLESIKN  
SECVHVSNEPMLATCVDSVGEIGVDVNSAASVSELPPEQEGALNSVPEISR  
GIPDVVSVGGGEENAAFNIOEIDSGDAAEVPLKSSVVEGREGNEAGSI  
VKREDFVAAEANTSGDRLKENKEVSESHSEVCEVNGVDSLNEENREVEQIV  
CGSNGGEFESDLSKKEKYVIEETTAQASVNAIEIPDKVEACVAGFETISSQ  
DKGIDESNGFLDEEPYKELOIGGAKDLTGDAKESVDTEDMDQVAKSKSEPK  
VDSSTTEIEITMRLEVDHVADEMDSKTVISSAVVTOFTGETSNDKETVMDVKEDVK  
DSEAGKSLDIHVPEATEBEDVDVAVGVGIEKGGVGGAEDAGQVLEIRENQEL  
SKEIAVDDETKISESEVYETIMIDEDQEKDNTDLAEVDENRDSVADIEGRSD  
HEDGVETQKETQETLVGAKTRKIAVSEETDTRDEDOEKDEMTVAEDVKRQDS  
SVADIERBESQEMTEQEDNSVMADEPEVEEENNSAGKRRKRGNTKVKGTGKK  
KEDYCFMCEFDGDLVLCDDR"  
2278..2365  
/rpt\_family="(GAA)n"  
2535..2614  
/note="exon predicted by xgrail, quality excellent"  
2940..3094  
/note="exon predicted by xgrail, quality good"  
complement(3316..3384)  
/note="exon predicted by xgrail, quality  
excellent\_shadowexon"  
3884..3920  
/note="exon predicted by xgrail, quality marginal"  
join(<3940..4068,4408..5287,5372..>5514)  
/gene="F16F14.2"  
<3940..>5514  
/gene="F16F14.2"  
/note="predicted by genscan and genefinder"  
join(3940..4068,4408..5287,5372..5514)  
/gene="F16F14.2"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="AAD22293.1"  
/db\_xref="PID:94544383"  
/db\_xref="GI:4544383"  
/translation="MESICGKCAQAVFPCIRGNKGICGTCMETVYKIERKOEKEPA  
QIDENDKTSWEILFDYVITDLKTQISPEPELDOAKRLKHEHNASQCASTEDVY  
IDGSGDLSVSPKRRTRSRKSGSEKILSSGDKNLSDETEHMASKEILLDVHMRG  
DSFPLPMEVQLLAYIKRYNLRPKRSGVTCDSRLQNLFGSHVGFEMNLNDS  
HFLKEQNQADQIDQIDIVTEPEPNHVDENDHFKVSGDKRKRTRKNVYKQSN  
LDDFAVAMHNINLYLRSLVEDLLESTAFEEKVASAFVRLRISGNOKODLYALVO  
VYGTSKAPEPVYVKKTTDYVLEILNDKTEVISIDISNDQFTEVTC"  
misc\_feature  
/note="exon predicted by xgrail, quality excellent"  
5693..5752  
misc\_feature  
/note="exon predicted by xgrail, quality excellent"  
5855..5905

```

misc_feature      6178..6260      /note="exon predicted by xgrall, quality excellent"
                    /note="exon predicted by xgrall, quality excellent"
mRNA              join(<6424..6583,6701..8419,8909..>9042)
                    /gene="F16F14.3"
gene              <6424..>9042
                    /gene="F16F14.3"
CDS               join(6424..6583,6701..8419,8909..9042)
                    /gene="F16F14.3"
                    /note="unknown protein"
                    /codon_start=1
                    /protein_id="A022314.1"
                    /db_xref="PID:94544404"
                    /db_xref="GI:4544404"
                    /translation="MULTIRECYEKLQLNLSPEERORRLIEPIHADPKKPDCESED
                    EDEKEKEKEKQLRPRSSSFNRRGRDIPISRKGFSSNESMTSTSYSPNSARELSRA
                    SYSGRSTGRGDYLGSSDDKVSDDMTSAREVOPSLGSEKRSYSIPETPARSELRA
                    IAPPELSPRIASEISMAPPAYVOPVPKSDSEKIWHYHDPGKVGKQVAFMAQDLRKWN
                    NNGYPAKLEIKRANESPLDSVILLTDALAGLPRKQYANDNSYMKAKQVAAFSQSSQS
                    ERLGFRARLAPTTITIPRNSODTWSGGSLPSPITNOQITTPAKRNFESRSPFKP
                    SPQSANQSMNVYVAQGOSQTSRIDI.PVVVNSGALQPTPIPTDPDINVSNSHAT
                    LHSPTPAGKRWSSGSMQTDHGSNTSPSSONNSYXGTPSPSYLPSOSOGFPGPDSMK
                    VAVPSQPNAAQACQMGNNMNNNNQNSAQPOAPANQNSMGOGVNNFMGMVPGQAQGV
                    NVNWGSSVPTVGTGTHSGWVA.PVQGGQAYPNPMGMPGPHGPOSSOQOYOQACTT
                    GSGMGPQGGIOGNSGNQNTQNTQNTAIPSGGQACGMKMGQSSQNDSDSGMNR
                    QSGLGNIQTDILKDPYPIGFIINPRLPECPPLVPLTQIHVNES"
repeat_region    6550..6583
                    /rpt_family="(GAA)n"
repeat_region    complement(8530..8554)
                    /rpt_family="(GA)n"
misc_feature      complement(9143..9189)
                    /note="exon predicted by xgrall, quality
                    marginal_shadowexon"
                    complement(9636..9699)
                    /rpt_family="AT_rich"
                    9914..9936
repeat_region    /note="exon predicted by xgrall, quality excellent"
                    join(10149..10485,10634..10765,10848..>10931)
                    /gene="F16F14.4"
mRNA              10149..>10931
                    /gene="F16F14.4"
gene              join(10306..10485,10634..10765,10848..10931)
                    /gene="F16F14.4"
CDS               /note="unknown protein"
                    /codon_start=1
                    /protein_id="A022294.1"
                    /db_xref="PID:94544384"
                    /db_xref="GI:4544384"
                    /translation="WLSKRLLALGNQSAITLAKPRGLSSGLDRPPFPKQSEEL
                    TANGRRATLVDTLALVRSLEAGVPSKQEAITSATELVNDLNSLEVSFSAKEM
                    KRLEMTQDSNLSFKSEKSSQVSTLVV"
misc_feature      11045..11130
                    /note="exon predicted by xgrall, quality good"
                    complement(11369..11430)
misc_feature      11442..11361
                    /note="exon predicted by xgrall, quality good_shadowexon"
                    complement(11625..11743)
misc_feature      /note="exon predicted by xgrall, quality good"
                    /note="exon predicted by xgrall, quality good_shadowexon"
                    11809..11875
misc_feature      /note="exon predicted by xgrall, quality good"
                    11966..12085
misc_feature      /note="exon predicted by xgrall, quality excellent"
                    complement(12034..12078)
misc_feature      /note="exon predicted by xgrall, quality
                    marginal_shadowexon"
                    <12986..>14269
mRNA              /gene="F16F14.5"
                    /gene="F16F14.5"
                    <12986..>14269
gene              /gene="F16F14.5"
                    /note="predicted by genscan and genefinder"
                    12986..14269
CDS               /gene="F16F14.5"

```

```

Query Match      9.9%; Score 18; DB 8; Length 114418;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY              10 CACATTCATTAAGTAAAC 27
Db 54374 CACATTCATTAAGTAAAC 54357
|||||

```

```

RESULT 10
AC002466
LOCUS           AC002466      78154 bp      DNA      PRI      20-AUG-1997
DEFINITION      Human BAC clone RG354L07 from 7q31, complete sequence.
ACCESSION       AC002466
NID             92337861
VERSION         AC002466.1 GI:2337861
KEYWORDS        HTG.
SOURCE           human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 78154)
AUTHORS         Miller, N and Gibson, A.
TITLE           The sequence of H. sapiens BAC clone RG354L07
JOURNAL         Unpublished (1997)
REFERENCE       2 (bases 1 to 78154)
AUTHORS         Waterston, R.
TITLE           Direct Submission
JOURNAL         Submitted (20-AUG-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT         SUBMITTED BY:
                Genome Sequencing Center
                Department of Genetics
                Washington University
                St. Louis MO 63108, USA
                http://genome.wustl.edu/gsc
                mailto:sapiens@watson.wustl.edu

```

NOTE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7> or send mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:  
This clone is from a release of the human BAC library. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: paeloBac11  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The actual start of this clone is at base position 1 of RG354L07; actual end is at 78154 of RG354L07. The orientation of this clone

is unknown.

This clone contains STS's SMS2071 (NID:91113356) and SMS2179 (NID:91113411).

## FEATURES

```

source          1..78154
                 Location/Qualifiers
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /chromosome="7"
                 /clone_lib="RG354107"
                 /clone_lib="CTTB-HS-A"
                 /map="7q31"
                 complement(575..827)
                 /rpt_family="ALU"
                 1985..2017
                 /rpt_family="L1"
                 3425..3629
                 /rpt_family="MER"
                 3721..3940
                 /rpt_family="MER"
                 complement(4181..4594)
                 /rpt_family="MER"
                 complement(4625..4923)
                 /rpt_family="L1"
                 complement(4939..5482)
                 /rpt_family="L1"
                 5991..6283
                 /rpt_family="ALU"
                 complement(6957..7181)
                 /rpt_family="MER"
                 complement(7430..7453)
                 /rpt_family="L1"
                 7957..7992
                 /rpt_family="L1"
                 10650..10692
                 /rpt_family="L1"
                 12746..12817
                 /rpt_family="L1"
                 13582..13599
                 /rpt_family="L1"
                 14112..14318
                 /rpt_family="L1"
                 16843..17210
                 /note="match to EST N73830 (NID:91211115) za56c01.s1"
                 /note="match to EST W00873 (NID:91272993) za56c01.r1"
                 20048..20337
                 /rpt_family="ALU"
                 20338..20367
                 /rpt_family="L1"
                 21568..21600
                 /rpt_family="L1"
                 22318..22608
                 /rpt_family="ALU"
                 25037..25083
                 /rpt_family="L1"
                 complement(28356..28928)
                 /rpt_family="L1"
                 33433..33459
                 /rpt_family="L1"
                 34370..34460
                 /rpt_family="L1"
                 34718..34737
                 /rpt_family="L1"
                 complement(35485..35775)
                 /rpt_family="ALU"
                 36208..37412
                 /rpt_family="L1"
                 complement(40472..40559)
                 /rpt_family="L1"
                 40482..40507
                 /rpt_family="L1"
                 40560..40854

```

```

repeat_region /rpt_family="ALU"
               complement(40874..42045)
repeat_region /rpt_family="L1"
               complement(45725..45762)
repeat_region /rpt_family="L1"
               47435..47710
repeat_region /rpt_family="ALU"
               48328..48357
repeat_region /rpt_family="L1"
               complement(49283..49415)
repeat_region /rpt_family="L1"
               complement(54609..54901)
repeat_region /rpt_family="ALU"
               complement(57040..57069)
repeat_region /rpt_family="L1"
               59551..59571
repeat_region /rpt_family="L1"
               59576..59705
repeat_region /rpt_family="ALU"
               complement(60076..60338)
repeat_region /rpt_family="ALU"
               complement(60918..60945)
repeat_region /rpt_family="L1"
               complement(60954..61254)
repeat_region /rpt_family="ALU"
               complement(62605..62633)
repeat_region /rpt_family="L1"
               63367..63437
repeat_region /rpt_family="L1"
               complement(63476..63766)
repeat_region /rpt_family="ALU"
               67301..67331
repeat_region /rpt_family="L1"
               complement(68442..68462)
repeat_region /rpt_family="L1"
               complement(73916..74206)
repeat_region /rpt_family="ALU"
               complement(76596..76628)
repeat_region /rpt_family="L1"

BASE COUNT  25369 a 13876 c 14122 g 24787 t
ORIGIN

```

Query Match 9.9%; Score 18; DB 11; Length 78154;  
 Best Local Similarity 100.0%; Pred.No. 11;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      86 TCTATATATGTTGCACAA 103
        |||
DB       4513 TCTATATGTTGCACAA 4530

```

```

RESULT  11
LOCUS    AC003087 103566 bp DNA PRI 06-NOV-1997
DEFINITION Human BAC clone RG119H12 from 7p15, complete sequence.
ACCESSION AC003087
NID       92588617
VERSION   AC003087.1 GI:2588617
KEYWORDS  HTG.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103566)
            Wollam,C and Edwards,J.
            The sequence of H. sapiens BAC clone RG119H12
            Unpublished (1997)
REFERENCE 2 (bases 1 to 103566)
            Waterston,R.
            Direct Submission
            Submitted (06-NOV-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

```

COMMENT

SUBMITTED BY:  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see  
http://www.nhgri.nih.gov/DIR/CTB/CHR7 or send  
mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:  
Clone RG119H12 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (http://www.resgen.com).  
VECTOR: pBelOBAC11  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The actual start of this clone is at base position 1 of RG119H12; actual end is at 103566 of RG119H12. The orientation of this clone is unknown.

This clone contains STS SWS61320 (NID:91222757).  
Location/Qualifiers  
1.103566  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/clone="RG119H12"  
/clone\_1lb="CTB-978SK-B"  
/map="7p15"  
complement(108..136)  
/rpt\_family="MER"  
complement(156..469)  
/rpt\_family="MER"  
863..1159  
/rpt\_family="ALU"  
2133..2423  
/rpt\_family="ALU"  
3048..3321  
/rpt\_family="MER"  
3322..3349  
/rpt\_family="LI"  
3372..3437  
/rpt\_family="MER"  
complement(3686..3976)  
/rpt\_family="ALU"  
complement(4045..4444)  
/rpt\_family="MER"  
complement(4465..4506)  
repeat\_region

FEATURES  
SOURCE

repeat\_region  
complement(4841..5133)  
/rpt\_family="ALU"  
complement(5313..5567)  
/rpt\_family="ALU"  
8285..8313  
/rpt\_family="LI"  
9053..9081  
/rpt\_family="LI"  
10769..10800  
/rpt\_family="LI"  
10801..11092  
/rpt\_family="ALU"  
11093..11464  
/rpt\_family="LI"  
complement(12740..12766)  
/rpt\_family="LI"  
complement(14279..14569)  
/rpt\_family="ALU"  
14727..14746  
/rpt\_family="LI"  
16277..16571  
/rpt\_family="ALU"  
complement(16720..16986)  
/rpt\_family="ALU"  
complement(17635..17742)  
/rpt\_family="LI"  
17969..18241  
/rpt\_family="ALU"  
18330..18381  
/rpt\_family="LI"  
19168..19512  
/rpt\_family="THE"  
19533..19742  
/rpt\_family="THR"  
19743..19767  
/rpt\_family="LI"  
21361..21382  
/rpt\_family="LI"  
complement(21900..22349)  
/rpt\_family="LI"  
complement(22451..22616)  
/rpt\_family="LI"  
24384..25389  
/rpt\_family="LI"  
complement(28788..29299)  
/rpt\_family="LI"  
29361..29435  
/rpt\_family="ALU"  
complement(29521..29621)  
/rpt\_family="LI"  
complement(29801..30161)  
/rpt\_family="LI"  
30163..30455  
/rpt\_family="ALU"  
complement(30515..31175)  
/rpt\_family="LI"  
complement(31189..31672)  
/rpt\_family="LI"  
complement(31773..32098)  
/rpt\_family="LI"  
32086..33943  
/rpt\_family="LI"  
34278..34340  
/rpt\_family="LI"  
complement(35983..35906)  
/rpt\_family="LI"  
complement(40028..40370)  
/rpt\_family="PRR"  
40046..40377  
/rpt\_family="PRR"  
41710..42086  
/note="similar to EST T59288 (NID:9661125) yc02c10.r1"  
misc\_feature

```

misc_feature      complement(42203..42360)
                  /note="similar to EST T91592 (NID:g723505) ye21h07.s1"
misc_feature      complement(42370..42449)
                  /note="similar to EST T91592 (NID:g723505) ye21h07.s1"
misc_feature      complement(43384..43677)
                  /note="similar to EST AA078512 (NID:g1837986)"
repeat_region     46506..47375
                  /rpt_family="PTR"
repeat_region     complement(46514..47369)
                  /rpt_family="PTR"
repeat_region     complement(47963..48263)
                  /rpt_family="L1"
repeat_region     complement(48288..48425)
                  /rpt_family="L1"
repeat_region     complement(48666..48698)
                  /rpt_family="L1"
repeat_region     complement(49025..50325)
                  /rpt_family="L1"
repeat_region     complement(50755..54270)
                  /rpt_family="L1"
repeat_region     complement(54271..54561)
                  /rpt_family="ALU"
repeat_region     complement(54562..57396)
                  /rpt_family="L1"
repeat_region     complement(57511..57714)
                  /rpt_family="L1"
repeat_region     complement(57720..57940)
                  /rpt_family="MER"
repeat_region     57942..58252
                  /rpt_family="ALU"
repeat_region     complement(58259..59082)
                  /rpt_family="MER"
repeat_region     59701..59841
                  /rpt_family="MER"
repeat_region     60644..61636
                  /rpt_family="PTR"

Query Match      9.9%; Score 18; DB 11; Length 103566;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TCTATATGTTTCACAA 103
    |||||
DB 3131 TCTATATGTTTCACAA 3114

RESULT 12
AC004911 155382 bp DNA PRI 28-DEC-1998
LOCUS Homo sapiens PAC clone D0870F17 from 7q33-q36, complete sequence.
AC004911
AC004911 93478665
NID 93478665
VERSION AC004911.1 GI:3478665
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155382)
AUTHORS Bauer,C., Langston,Y., Harrison,M. and Lennox,S.
TITLE The sequence of Homo sapiens PAC clone D0870F17
JOURNAL Unpublished (1998)
AUTHORS 2 (bases 1 to 155382)
TITLE Waterston,R.H.
JOURNAL Direct Submission
          Submitted (12-JUN-1998) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 3 (bases 1 to 155382)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1998) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Aug 27, 1998 this sequence version replaced g1:3213086.

```

SUBMITTED BY: WUGSC  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:saplens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see  
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send  
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. For further details, see http://bacpac.med.buffalo.edu/  
The clone is available from Genome Systems, Inc.  
(http://www.genomesystems.com).

VECTOR: pcypac2  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is DJ0981B16, the clone sequenced to the right is DJ1159C10. Actual start of this clone is at base position 1 of DJ0870F17; actual end is at 155382 of DJ0870F17.

## FEATURES

```

SOURCE
1..155382
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="7"
   /map="7q33-q36"
   /clone_1lb="RPCI-5"
   /clone="DJ0870F17"
1..361
   /rpt_family="L1"
repeat_region 803..1029
   /rpt_family="Alu"
repeat_region 2035..2204
   /rpt_family="L1"
repeat_region 2210..2377
   /rpt_family="L1"
repeat_region 2378..2791
   /rpt_family="WALR"
repeat_region 2792..2961
   /rpt_family="L1"
repeat_region 3050..3465
   /rpt_family="L1"
repeat_region 3471..3858
   /rpt_family="L1"
repeat_region 3932..4937
   /rpt_family="L1"
repeat_region 5732..6026
   /rpt_family="Alu"
repeat_region 7143..7449
   /rpt_family="Alu"

```

```

repeat_region      7705..7939
                    /rpt_family="L2"
repeat_region      8442..8513
                    /rpt_family="MER2_type"
repeat_region      8550..8605
                    /rpt_family="MER2_type"
repeat_region      8618..9335
                    /rpt_family="MER2L_g"
repeat_region      9384..9502
                    /rpt_family="MALR"
repeat_region      9869..9931
                    /rpt_family="Retroviral"
repeat_region      9979..10035
                    /rpt_family="MALR"
repeat_region      10286..10592
                    /rpt_family="Alu"
                    13148..13222
misc_feature       /note="match to EST AA140531 (NID:g1700092) mq03g11.r1"
                    complement(13150..13210)
misc_feature       /note="match to EST A1020949 (NID:g3235285) ub06g06.r1"
                    complement(13155..13212)
misc_feature       /note="match to EST A1020949 (NID:g3235285) ub06g06.r1"
                    13156..13201
                    /note="match to EST AA261046 (NID:g1897566) mz81a02.r1"
                    complement(13157..13212)
misc_feature       /note="match to EST AA818166 (NID:g2888046)"
                    complement(13160..13211)
misc_feature       /note="match to EST AA818166 (NID:g2888046)"
                    complement(13164..13210)
misc_feature       /note="match to EST AA128837 (NID:g1688718) z010g09.r1"
                    13166..13209
                    /note="match to EST AA684287 (NID:g2670873) vm54b03.r1"
                    13167..13210
                    /note="match to EST AA261046 (NID:g1897566) mz81a02.r1"
                    13167..13212
misc_feature       /note="match to EST AA140531 (NID:g1700092) mq03g11.r1"
                    13167..13212
misc_feature       /note="match to EST AA684287 (NID:g2670873) vm54b03.r1"
                    13169..13209
                    /note="match to EST AA247116 (NID:g1878557)"
                    complement(13829..13845)
misc_feature       /note="similar to EST AA134277 (NID:g1691756) z006a09.r1"
                    13841..13987
                    /rpt_family="L1"
repeat_region      14055..14342
                    /rpt_family="L1"
repeat_region      14936..15084
                    /rpt_family="L1"
repeat_region      15262..15562
                    /rpt_family="Alu"
                    15629..16383
repeat_region      /rpt_family="L1"
                    16400..16922
repeat_region      /rpt_family="L1"
                    16923..17088
                    /rpt_family="L1"
repeat_region      17089..17442
                    /rpt_family="L1"
repeat_region      17528..17954
                    /rpt_family="MALR"
                    17992..18538
repeat_region      /rpt_family="L1"
                    18539..18922
repeat_region      /rpt_family="Retroviral"
                    18924..20551
repeat_region      /rpt_family="Retroviral"
                    20552..20814
repeat_region      /rpt_family="Retroviral"
                    20815..20920
repeat_region      /rpt_family="L1"
                    21186..21263
repeat_region      /rpt_family="L1"
                    21264..21751

```

```

repeat_region      /rpt_family="Retroviral"
                    21756..21940
                    /rpt_family="L1"
repeat_region      21950..22758
                    /rpt_family="L1"
repeat_region      22816..22957
                    /rpt_family="L1"
repeat_region      23094..23180
                    /rpt_family="MER2_type"
                    23174..23232
repeat_region      /rpt_family="MER2_type"
                    23237..23368
repeat_region      /rpt_family="MALR"
                    23365..23562
repeat_region      /rpt_family="L1"
                    23563..23774
repeat_region      /rpt_family="MALR"
                    23780..23953
repeat_region      /rpt_family="MER2_type"
                    25665..25959
repeat_region      /rpt_family="Alu"
                    26130..26589
repeat_region      /rpt_family="L1"
                    26590..26901
repeat_region      /rpt_family="L1"
                    27110..27385
repeat_region      /rpt_family="MER4_group"
                    28458..28574

Query Match
Best Local Similarity 9.9%; Score 18; DB 11; Length 155382;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 AACACGCAAAATCAAAA 178
Db 124015 AACACGCAAAATCAAAA 124032
|||||
|||||

RESULT 13
LOCUS HSCOS10 40247 bp DNA PRI 06-DEC-1998
DEFINITION Homo sapiens chromosome 3 subtelomeric region.
ACCESSION AF109718
VERSION 93978547
KEYWORDS AF109718.1 GI:3978547
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 40247)
Traak,B.J., Rowen,L., Nguyen,O., Massa,H.F., Giorgi,D.,
Brand-Arpon,V. and Rouquier,S.
Sequence of a genomic duplication found at the subtelomeric region
of several chromosomes
Unpublished
2 (bases 1 to 40247)
Rowen,L.
Direct Submission
Submitted (27-NOV-1998) Department of Molecular Biotechnology, Box
357730 University of Washington, Seattle, Washington 98195, USA
Sequencing methodology: High redundancy shotgun. Interspersed
Repeats were identified with RepeatMasker (available from
http://ftp.genome.washington.edu/RW/RepeatMasker.html).
FEATURES
source
1..40247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="6P21"
/clone_lib="LLO3NC01"
674..816
/rpt_family="(TAA)n"

```

repeat\_region 876..1076  
/rpt\_family="L1MA5"  
repeat\_region 1073..1475  
/rpt\_family="L191\_5end"  
repeat\_region 1078..1836  
/rpt\_family="L143\_5end"  
repeat\_region complement(1974..2002)  
/rpt\_family="AT\_rich"  
repeat\_region complement(2390..2489)  
/rpt\_family="LINE2"  
repeat\_region complement(2624..2893)  
/rpt\_family="Alusx"  
repeat\_region complement(3330..3426)  
/rpt\_family="LINE2"  
repeat\_region complement(3438..3735)  
/rpt\_family="Alusg"  
repeat\_region complement(4234..4517)  
/rpt\_family="Alusx"  
repeat\_region complement(4520..4630)  
/rpt\_family="FLAM\_C"  
repeat\_region 4631..4733  
/rpt\_family="FLAM\_C"  
repeat\_region 4841..5146  
/rpt\_family="Alusg"  
repeat\_region 5388..5590  
/rpt\_family="L191\_5end"  
repeat\_region 5598..5742  
/rpt\_family="Alusx"  
repeat\_region 5745..5818  
/rpt\_family="Alusg"  
repeat\_region complement(6298..6547)  
/note="96% nt similarity to chimp telomeric DNA clone  
pcht9 in X74283"  
misc\_feature 6389..8826  
/note="Similarity to human mRNA for KIAA0306 gene in  
AB002304"  
misc\_feature 7862..8159  
/note="98% nt similarity to H.sapiens telomeric DNA clone  
160TEL022 in 296324"  
repeat\_region complement(8921..9220)  
/rpt\_family="Alusg"  
misc\_feature 9789..10238  
/note="match: 3' ESTs AI200936, AI015622, AI223153  
Soares-testis cDNA"  
misc\_feature 11047..13152  
/note="48-base repeat region"  
repeat\_region complement(14390..14809)  
/rpt\_family="MER1A"  
repeat\_region 15232..15515  
/rpt\_family="Alusg"  
repeat\_region complement(15764..15872)  
/rpt\_family="Alusg"  
misc\_feature 15880..16108  
/note="similarity to H.sapiens telomeric DNA sequence  
7PTEL027 in 296680"  
repeat\_region complement(16109..16408)  
/rpt\_family="Alusg"  
repeat\_region 16520..16592  
/rpt\_family="MER2"  
repeat\_region 16622..16808  
/rpt\_family="Alusg"  
repeat\_region 17415..17549  
/rpt\_family="Alusg"  
repeat\_region complement(17804..18223)  
/rpt\_family="MER1A"  
repeat\_region 18506..18591  
/rpt\_family="FLAM\_A"  
repeat\_region 18591..18662  
/rpt\_family="Alusg"  
repeat\_region 18672..18880  
/rpt\_family="Alusg"  
repeat\_region complement(18974..19359)  
/rpt\_family="MER21B"

repeat\_region complement(19385..19693)  
/rpt\_family="Alusg"  
repeat\_region 19764..20066  
/rpt\_family="Alusg"  
repeat\_region complement(20068..20310)  
/rpt\_family="MER21B"  
repeat\_region 20475..20586  
/rpt\_family="MER21B"  
repeat\_region complement(21267..21368)  
/rpt\_family="Alusg"  
repeat\_region 21402..21531  
/rpt\_family="Alusg"  
repeat\_region complement(21540..21971)  
/rpt\_family="MER1A"  
repeat\_region 22331..22621  
/rpt\_family="Alusg"  
repeat\_region 22624..22698  
/rpt\_family="Alusg"  
repeat\_region 22714..23008  
/rpt\_family="Alusg"  
repeat\_region 23095..23401  
/rpt\_family="Alusg"  
repeat\_region 23423..23574  
/rpt\_family="LINE2"  
repeat\_region 23625..23985  
/rpt\_family="LINE2"  
repeat\_region 24177..24330  
/rpt\_family="MER1D"  
repeat\_region 24337..24630  
/rpt\_family="Alusg"  
repeat\_region 24884..25155  
/rpt\_family="Alusg"  
repeat\_region complement(25168..25269)  
/rpt\_family="Alusg"  
repeat\_region 25270..25565  
/rpt\_family="Alusg"  
repeat\_region complement(25588..25643)  
/rpt\_family="Alusg"  
repeat\_region 25734..26025  
/rpt\_family="Alusg"  
repeat\_region 26026..26099  
/rpt\_family="L1MB6"  
repeat\_region complement(26148..26258)  
/rpt\_family="FLAM\_A"  
repeat\_region 26263..26876  
/rpt\_family="L1MB8"  
repeat\_region 26900..27201  
/rpt\_family="Alusg"  
repeat\_region 27504..27807  
/rpt\_family="Alusg"  
repeat\_region complement(28258..28557)  
/rpt\_family="Alusg"  
repeat\_region complement(28585..28717)  
/rpt\_family="FLAM\_C"  
repeat\_region 28817..29421  
/rpt\_family="LINE2"  
repeat\_region complement(29422..29584)  
/rpt\_family="Alusg"  
repeat\_region complement(29585..29872)  
/rpt\_family="Alusg"  
repeat\_region complement(29885..30000)  
/rpt\_family="FLAM\_C"  
repeat\_region complement(30021..30378)  
/rpt\_family="MER1D"  
repeat\_region 30382..30672  
/rpt\_family="Alusg"  
repeat\_region complement(30676..30771)  
/rpt\_family="MER21B"  
repeat\_region complement(30766..30926)  
/rpt\_family="MER21B"  
repeat\_region complement(30932..31225)  
/rpt\_family="Alusg"  
repeat\_region complement(31238..31491)



```

repeat_region /rpt_family="MER21B"
31536..31840 /rpt_family="AluJo"
repeat_region complement(31844..32116)
/rpt_family="MER21B"
repeat_region 32289..33205
/rpt_family="MER11A"
repeat_region 32586..33205
/rpt_family="MER11B"
repeat_region 33333..33440
/rpt_family="MIR"
repeat_region complement(34103..34215)
/rpt_family="AluDb"
repeat_region complement(34223..34523)
/rpt_family="AluJo"
repeat_region 34562..34659
/rpt_family="LIMC4"
repeat_region 34850..34931
/rpt_family="TAAA)n"
repeat_region 34934..35228
/rpt_family="AluJo"

```

```

Query Match 9.9%; Score 18; DB 11; Length 40247;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 133 AGAACAAACACACACAA 150
Db 15507 AGAACAAACACACACAA 15524

```

```

RESULT 14
LOCUS AB017156 2937 bp mRNA ROD 04-MAR-1999
DEFINITION Mus musculus gob-5 mRNA, complete cds.
ACCESSION AB017156
NID 93721911
VERSION AB017156.1 GI:3721911
KEYWORDS GOB-5.
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Komiya,T., Tanigawa,Y. and Hirohashi,S.
TITLE Cloning and identification of the gene gob-5, which is expressed in
JOURNAL intestinal goblet cells in mice
MEDLINE Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)
99160866
2 (bases 1 to 2937)
Komiya,T., Tanigawa,Y. and Hirohashi,S.
DIRECT SUBMISSION Submitted (26-AUG-1998) to the DDBJ/EMBL/Genbank databases. Tohru
Komiya, ERATO, JST, Hirohashi Cell Configuration Project; 5-9-9,
Tokodai, Tsukuba, Ibaraki 300-2635, Japan
(E-mail:tkom@hccp.jst.go.jp, Tel:81-298-47-7563,
Fax:81-298-47-5226)
FEATURES
source location/Qualifiers
1..2937
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="goblet cell"
/dev_stage="adult"
/tissue_type="intestine"
15..2756
/gene="gob-5"
15..2756
/gene="gob-5"
15..2756
/gene="gob-5"
/codon_start=1
/protein_id="BAA33743.1"
/db_xref="PID:d1034712"
/db_xref="PID:93721912"
/db_xref="GI:3721912"

```

```

/translation="MESLSPVFLILHLEGLSRLIQNNNGEYVIAIDHPV
EDBALQHTKDWTOASPYLEKATGRFTFKNAIILPESMKAPETTPKLTENNA
DVLVSTTSPLENDPEPTERHIGACGKGIHLLPDLAKKLQYGPORDTFVEMAH
FRWGVEYENENDEKFTLSKGPQAVKCSAIIKGNVRKCGGSCITNCKVDRYTG
LYKDNCEVFEVDPHONKASIMFNONINSVECTEKNNHNOEAPNDONCINLSTEV
IOESSEDFKOTMTAOPAPPTSLIOGRIQVCLVDRSGSMILRMOASRLFL
LIQTVRQGSWGVGVTDSAAVYQSEIKQNSGADRDLRLKHLPTVAGSTSTCSGRT
AFVIRKATPTDSEIVLITDEGNTISCFDLVKOSGAIHRTVALGPAAKELBOL
KMTGLQTYSDQVONNGLVDAFALSSGNAALIAOHSIQLESKVNQNNWNGSVI
VDSVSGKDLFLITWTTHPTIFIMDPSGEONGFIIDTTKVAAYLVPGTAKVGFK
VSIQASSOTLITVTSRASATLPITVPPVYKNGKFPSPVYVASIROGASPIIR
ASVTLIESVNGKVTLELDNGAGADARKNGVYSRFTDANGRSYKIMAGGV
TSDROAAPKPRAMTIDGWIEDGFRNMPRPRTSYVDKOLCFSTSSGSEFVATN
VPAAPIDPLFPQCITDLKASIQGNLVNLWTAGADYDHRASNTIIRKSTSLVD
LRDHFNTSLQVNTGTGLIPKASEEIEFEELGNTFGNTDITFIALQAVDKSLKSEI
SNIRAVSVFIPAOEPIPDSTPCPDISINSTIPIQIHVKIMMKVLGEMOYVLGLH"
BASE COUNT 860 a 718 c 693 g 666 t
ORIGIN

```

```

Query Match 9.9%; Score 18; DB 12; Length 2937;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 119 GTTGAAATCTGTACAGAA 136
Db 756 GTTGAAATCTGTACAGAA 773

```

```

RESULT 15
LOCUS AC004676 201175 bp DNA HTG 10-MAY-1998
DEFINITION Homo sapiens chromosome X clone bwxdl71, WORKING DRAFT SEQUENCE, 1
ordered pieces.
ACCESSION AC004676
NID 93126882
VERSION AC004676.1 GI:3126882
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 201175)
AUTHORS Chen,E., Brownstein,B.H., States,D.J., Schlessinger,D. and
Mazzarella,R.
TITLE Direct Submission
JOURNAL Unpublished (1997)
2 (bases 1 to 201175)
BROWNSTEIN,B.H., STATES,D.J. and MAZZARELLA,R.
DIRECT SUBMISSION Submitted (10-MAY-1998) Center for Genetics in Medicine, Box 8232,
Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA
Current status of this project is available at:
'http://www.lbc.wustl.edu/cgm/seq_projects.html'
Submitted by:
Elison Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp.,
850 Lincoln Center Drive,
Foster City, CA 94404 USA
e-mail: elison@genseq.apdbio.com
and

```

Buddy Brownstein,  
Center for Genetics in Medicine,  
Washington University School of Medicine, Box 8232  
4566 Scott Avenue,  
St. Louis, MO 63110, USA  
e-mail: buddy@genetics.wustl.edu  
and

David J. States,  
Institute for Biomedical Computing  
Washington University in St. Louis  
700 South Euclid Ave.  
St. Louis, MO 63108 USA  
e-mail: states@dbc.wustl.edu.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved  
1 201175: contig of 201175 bp in length.

# FEATURES

Location/Qualifiers  
1..201175  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DMD171"  
/chromosome="X"

BASE COUNT 60554 a 41437 c 41653 g 57531 t  
ORIGIN

Query Match 9.98; Score 18; DB 34; Length 201175;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AACACAGCAAAATCAAAA 177  
|||||  
Db 192394 AACACAGCAAAATCAAAA 192377

Search completed: August 6, 1999, 11:33:59  
Job time: 6125 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 1999, 11:32:51 ; Search time 3266.17 Seconds

(without alignments)  
265.661 Million cell updates/sec

Title: US-09-049-696-2

Perfect score: 273  
Sequence: 1 GTTGCAATCGACCCCAATGT.....AGATGGCAACTGTGGCAG 273

Scoring table: OLIGO\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBdb1: \*  
1: gb\_dal: \*  
2: gb\_dal: \*  
3: gb\_ov: \*  
4: gb\_ov: \*  
5: gb\_ov: \*  
6: gb\_ov: \*  
7: gb\_ov: \*  
8: gb\_ov: \*  
9: gb\_ov: \*  
10: gb\_ov: \*  
11: gb\_ov: \*  
12: gb\_ov: \*  
13: gb\_ov: \*  
14: gb\_ov: \*  
15: gb\_ov: \*  
16: gb\_ov: \*  
17: gb\_ov: \*  
18: gb\_ov: \*  
19: gb\_ov: \*  
20: gb\_ov: \*  
21: gb\_ov: \*  
22: gb\_ov: \*  
23: gb\_ov: \*  
24: gb\_ov: \*  
25: gb\_ov: \*  
26: gb\_ov: \*  
27: gb\_ov: \*  
28: gb\_ov: \*  
29: gb\_ov: \*  
30: gb\_ov: \*  
31: gb\_ov: \*  
32: gb\_ov: \*  
33: gb\_ov: \*  
34: gb\_ov: \*  
35: gb\_ov: \*  
36: gb\_ov: \*  
37: gb\_ov: \*  
38: gb\_ov: \*  
39: gb\_ov: \*  
40: gb\_ov: \*  
41: gb\_ov: \*  
42: gb\_ov: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	169	61.9	3311	11	AF039400 Homo sapi

Result	Score	Match	Length	ID	Description
169	61.9	2826	42	AF127036	AF127036 Homo sapi
105	38.5	35278	11	AF039401	AF039401 Homo sapi
26	9.5	2937	12	AB017156	AB017156 Mus muscu
22	8.1	16780	11	H0091323	H0091323 Human Chrom
19	7.0	180388	9	H0091323	H0091323 Human Chrom
19	7.0	135928	10	HS434014	HS434014 Homo sapi
8	7.0	320	14	G48840	G48840 SHGC-83613
9	7.0	190000	34	AC004479	AC004479 Homo sapi
10	7.0	306131	34	AC006874	AC006874 Caenorhab
11	7.0	110000	34	CEY105C5_5	Continuation (6 of
12	7.0	180821	35	AC007115	AC007115 Homo sapi
13	7.0	55629	35	AC007593	AC007593 Drosophi
14	7.0	163362	35	AC007624	AC007624 Drosophi
15	7.0	40014	36	CELF10C1	CELF10C1
16	6.6	2937	1	HGEMR	HGEMR
17	6.6	2938	1	HGEMR	HGEMR
18	6.6	60232	2	AE001272	AE001272 Lactococc
19	6.6	900	5	E03986	E03986 DNA encodin
20	6.6	2492	7	PUR28645	PUR28645 phaseolus v
21	6.6	37947	8	SCU39205	SCU39205 Saccharomyc
22	6.6	3259	8	SCU50630	SCU50630 Saccharomyc
23	6.6	90169	9	HS398A12	HS398A12 Human DNA
24	6.6	119198	9	HS845024	HS845024 Homo sapi
25	6.6	107761	10	HS417022	HS417022 Homo sapi
26	6.6	114638	10	HS799N4	HS799N4 Human DNA
27	6.6	44057	10	HSCN37F10	HSCN37F10 Human DNA
28	6.6	76782	10	HSU501W23	HSU501W23 Human DNA
29	6.6	175839	11	AC004491	AC004491 Homo sapi
30	6.6	150332	11	AC004921	AC004921 Homo sapi
31	6.6	186986	11	AC006210	AC006210 Homo sapi
32	6.6	2088	11	HSU77735	HSU77735 Human plm-2
33	6.6	77197	11	U90093	U90093 Human Chrom
34	6.6	884	12	GP1GMBP1	GP1GMBP1 Cavia cobay
35	6.6	36675	34	AC006671	AC006671 Caenorhab
36	6.6	262336	34	AC006786	AC006786 Caenorhab
37	6.6	262336	34	AC006786	AC006786 Caenorhab
38	6.6	207370	34	AC006798	AC006798 Caenorhab
39	6.6	342379	34	CEY39B6	CEY39B6 Caenorhabd
40	6.6	116846	34	HS191E19	HS191E19 Homo sapi
41	6.6	116271	34	HS39417	HS39417 Homo sapi
42	6.6	157221	34	HS619N21	HS619N21 Homo sapi
43	6.6	185092	34	HS761G12	HS761G12 Homo sapi
44	6.6	134899	35	AC006571	AC006571 Homo sapi
45	6.6	349657	35	AC007351	AC007351 Homo sapi

## ALIGNMENTS

Result	Score	Match	Length	ID	Description
1	169	61.9	3311	11	AF039400 Homo sapi
AF039400	3311 bp	MRNA	15-DEC-1998		
LOCUS					
DEFINITION					
complete cds.					
ACCESSION					
AF039400					
VERSION					
94009457					
KEYWORDS					
AF039400.1					
SOURCE					
ORGANISM					
human.					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
Eutheria; Primates; Catarrhini; Homiidae; Homo.					
1 (bases 1 to 3311)					
Gruber/A.D., Elble/R.C., Ji/H.L., Schreuer/K.D., Fuller,C.M. and					
Pauli,B.U.					
Genomic cloning, molecular characterization, and functional					
analysis of human CCAL, the first human member of the family of					
Ca2+-activated Cl- channel proteins					
Genomics 54 (2), 200-214 (1998)					
99047526					
2 (bases 1 to 3311)					
Gruber/A.D., Elble,R. and Pauli,B.U.					
Direct Submission					

```
JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,  
Cornell University, Ithaca, NY 14853, USA  
FEATURES Location/Qualifiers  
SOURCE 1. .3311
```

```

gene      1. .3311
          /gene="hclCA1"
CDS       352. .3096
          /CDS="hclCA1"

```

BASE COUNT	1028 a	692 c	742 g	849 t
ORIGIN				

Query Match	61.98;	Score 169;	DB 11;	Length 3311;
Best Local Similarity	99.5%;	Pred. No. 5.6e-86;		
Matches 219;	Conservative	0;	Mismatches 1;	Indels 0;
			Gaps	0;

QY 1 GTTGCATTCGACCCCAATGTCGCCAGGAGATGAAGAACATCAATTCACCAAAATTAAGGCATG 60

Db 460 GTTGAATTCGACCCCAATGTCGCCAGGAGATGAAGAACATCAATTCACCAAAATTAAGGCATG 519

QY 61 GTGACCCAGGCACTCTGTATCTGTTTGAAGCTACAGGAAAGCATTTTATTTCAAAAT 120

Db 520 GTGACCCAGGCACTCTGTATCTGTTTGAAGCTACAGGAAAGCATTTTATTTCAAAAT 579

QY 121 GTTGCATTTTGATTCCTGGAACATGGAAGACAAAGGTGACTATGTGAGACCCAAACTT 180

Db 580 GTTGCATTTTGATTCCTGGAACATGGAAGACAAAGGTGACTATGTGAGACCCAAACTT 639

QY 181 GAGACTACAAAATGCTGATGTTCTGTTCTGAGTCA 220

Db 640 GAGACTACAAAATGCTGATGTTCTGTTCTGAGTCA 679

RESULT	2		
AF127036			
LOCUS	AF127036	2826 bp	mRNA
DEFINITION	Human <i>sepiens</i> calcium-activated chloride channel protein 1 (CaCC1		PR1 16-APR-1999
ACCESSION	AF127036	mRNA, complete cds.	
NID	94585468		
VERSION	AF127036.1	GI:4585468	
KEYWORDS			
SOURCE	human.		
ORGANISM	*Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		

REFERENCE  
AUTHORS  
TITLE  
channel

1 (bases 1 to 2826)  
Epithelia; Primates; Catarrhini; Hominoidea; Homo.  
Agne, M., Vernat, T. and Cloussou, J.-M.  
Cloning of three human homologs of bovine epithelial chloride  
channel

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2826)
AUTHORS	Agnel, M. and Culouscou, J.-M.
TITLE	Direct Substitution
JOURNAL	Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des
FEATURES	Carrieres, Rueil-Malmaison 92500, France
SOURCE	Location/Qualifiers
	1..2826

```
gene      1. .2826  
/gene="CaCC1"  
CDS      5. .2749  
/gene="CaCC1"
```

BASE COUNT	875 a	623 c	632 g	696 t
ORIGIN				

Query Match	61.9%	Score 169;	DB 42;	Length 2826;
Best Local Similarity	99.5%;	Pred. NO. 5.5e-86;		
Matches 219;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

QY	1	GTTCGATCCGACCCCAATGTGCAGAAAGATGAAACATCATTTCAACAAATAAAGACATG	60
Db	113	GTTCGAAATCGACCCCAATGTGCAGAAAGATGAAACATCATTTCAACAAATAAAGACATG	1
QY	61	GTGACCCGAGCATCTCTGTATCTGTTTGAAGCTACAGGAAAGCATTTTATTTCAAAAT	120
Db	173	GTGACCCGAGCATCTCTGTATCTGTTTGAAGCTACAGGAAAGCATTTTATTTCAAAAT	2
QY	121	GTTCGCAATTTGATTCCTGTAACATGGAAGACAAAGGNTGACTATGTGAGCCAAAATT	180
Db	233	GTTCGCAATTTGATTCCTGTAACATGGAAGACAAAGGCTGACTATGTGAGCCAAAATT	23
QY	181	GAGACCTACAAAATGCTGATGTTCGTTGCTGAGATCTA	220
Db	293	GAGACCTACAAAATGCTGATGTTCGTTGCTGAGATCTA	332

RESULT	3				
AF039401					
LOCUS	AF039401	35278	bp	DNA	
DEFINITION	Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) gene complete cds.				
					15-DEC-1998
					PR1

ACCESSION AF039401  
NID 94009459  
VERSION AF039401.1 GI:4009459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Eblie,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and Pauli,B.U.  
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins  
JOURNAL Genomics 54 (2), 200-214 (1998)  
MEDLINE 99047526  
REFERENCE 2 (bases 1 to 35278)  
AUTHORS Gruber,A.D., Eblie,R. and Pauli,B.U.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1997) Department of Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA  
FEATURES  
source 1..35278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22-p31"  
1..1617  
/gene="hCLCA1"  
1..33522  
/gene="hCLCA1"  
767..1265  
/gene="hCLCA1"  
/rpl\_family="L1 transposable element"  
/rpl\_type="dispersed"  
1587..1596  
/gene="hCLCA1"  
/note="putative"  
1618..1708  
/gene="hCLCA1"  
/number=1  
join(1618..1708,1962..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..22427,24522..24628,26644..26859,27447..27708,28766..28936,31840..32079,32919..33522)  
/gene="hCLCA1"  
/product="calcium-dependent chloride channel-1"  
1962..2383  
/gene="hCLCA1"  
/number=2  
join(2222..2383,6667..6807,6979..7126,9696..9801,15454..15631,18591..18809,19783..20010,22253..24522,24628,26644..26859,27447..27708,28766,31840..32079,32919..33310)  
/gene="hCLCA1"  
/codon\_start=1  
/product="calcium-dependent chloride channel-1"  
/protein\_id="AAC95429.1"  
/db\_xref="PID:g4009460"  
/db\_xref="GI:4009460"  
translation="MGPEKSSVFLLHLLEGALSNSLIQNNNGYGIIVAIIDPNV  
EDFTLQIKDVTALVLEFATGKRFYRNVALIPETWTKRADVRSVRLKRYKA  
DIVVABSTPGNDPEPTQKNGCEKERRHLPPDFAGKLAIEYQGRAPFHEAH  
LRNGVEPTNDEFTYLSNGRIQVRSASITGNVVKCKGGSCITKRTFNKVTGL  
YERGECEVLQSRQTERKASIMEAQHVDIVFCEQNNKKAQPNKQNCMLRSTWEI  
RDESEFRKTPMTOPNPPTSLIQIGRTICVLDKSGSMATGNRLNNOAGOLEI  
LQVELGSVGMATFDSAAHOSLIIQINGSDRDTAKRLPAASGSGTICGLRBA  
FYIIRKYPDGSSEIVLTDGENTISGCNEVKSQALIHVTAIGPSAODEEISK  
MTGGLOIYADQVQNNGLIDAFALSSGNAYVRSRQLSKGLTQNSGMNNTYIV  
DSTVGKDTLITWITPOIILMDPSGQGGGVNKKMAIYLQIPGAKGTMY  
SLQASSOTLITVTSRASNTLPPITVTSNTKDTSKFSPPLVYVYAIIRGASPIILNA  
SVATLIESVNGKVTLELDNGAGADATKDGYSRTFTYDNGRYSVYRVLGYN  
AARRRVIPQSGALYIPGMIENDEIQNPPREINRNDVDHKKVCFERTSSGSEFVS  
DVNAPIPDLPFPQGITDAAEIHGSLINLWTAPEDDDYDHGAHAIIRISTIID

LRDKNESLQVNTTALIPKPEANSEEVLFKPENTIFENGTLFLIAQVADKVKSEI  
SNIAKVSLEFIPQIPETPSPDETSAPCPVHINSTIPGIHLIKIMKMWIGELQLSIA  
exon 6667..6807  
/gene="hCLCA1"  
/number=3  
exon 6979..7126  
/gene="hCLCA1"  
/number=4  
exon 9696..9801  
/gene="hCLCA1"  
/number=5  
exon 15454..15631  
/gene="hCLCA1"  
/number=6  
exon 18591..18809  
/gene="hCLCA1"  
/number=7  
exon 19783..20010  
/gene="hCLCA1"  
/number=8  
exon 22253..22427  
/gene="hCLCA1"  
/number=9  
exon 24522..24628  
/gene="hCLCA1"  
/number=10  
exon 26644..26859  
/gene="hCLCA1"  
/number=11  
exon 27447..27708  
/gene="hCLCA1"  
/number=12  
exon 28766..28936  
/gene="hCLCA1"  
/number=13  
exon 31840..32079  
/gene="hCLCA1"  
/number=14  
exon 32919..33522  
/gene="hCLCA1"  
/number=15  
BASE COUNT 11245 a 6870 c 7230 g 9933 t  
ORIGIN  
Query Match 38.5%; Score 105; DB 11; Length 35278;  
Best Local Similarity 100.0%; Pred. No. 2,2e-49;  
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 53 AGACATGCTGACCCAGCATCTCTGATCTGTTGAGCTACGAAAGCATTTATT 112  
|||||  
Db 6665 AGGACATGCTGACCCAGCATCTCTGATCTGTTGAGCTACGAAAGCATTTATT 6724  
QY 113 TCAAAATGTTGCCATTTGATTCCTGAACATGAGACAAAG 157  
|||||  
Db 6725 TCAAAATGTTGCCATTTGATTCCTGAACATGAGACAAAG 6769  
RESULT 4  
LOCUS AB017156 2937 bp mRNA ROD 04-MAR-1999  
DEFINITION Mus musculus gob-5 mRNA, complete cds.  
ACCESSION AB017156  
NID 93721911  
VERSION AB017156.1 GI:3721911  
KEYWORDS GOB-5;  
SOURCE Mus musculus adult intestine goblet cell cDNA to mRNA.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 2937)  
AUTHORS Komiyama,T., Tanigawa,Y. and Hirohashi,S.

**TITLE** Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice

**JOURNAL** Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)

**MEDLINE** 99160866

**REFERENCE** 2 (bases 1 to 2937)

**AUTHORS** Komiyama, T., Tanigawa, Y. and Hirohashi, S.

**TITLE** Direct Submission

**JOURNAL** Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru Komiyama, ERATO, JST, Hirohashi Cell Configuration Project; 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan

(E-mail: tkom@nccp.jst.go.jp, Tel: 81-298-47-7563, Fax: 81-298-47-5226)

**FEATURES** Location/Qualifiers

**source** 1. 2937

**organism** "Mus musculus"

**/db\_xref="taxon:10090"**

**/cell\_type="goblet cell"**

**/dev\_stage="adult"**

**/tissue\_type="intestine"**

**/gene="gob-5"**

**/gene="gob-5"**

**/codon\_start=1**

**/protein\_id="BA033743.1"**

**/db\_xref="PIR:G1034712"**

**/db\_xref="PIR:G1034712"**

**/db\_xref="GI:3721912"**

**/translation="MESLSPVPLILHLLEGLVSESLIQLNNNGEGIVIAIDHVP  
EDALIOHKDWTQAPYLFEGATGRFEKFNVAIIIPESMKAPETPTPKLETFENA  
DVIVSTSPGNDPEYTHIEGEGEITLHLPDLAKKLQVPODRTVHEHAH  
FRKGVNEYNNDEKFLSKGRPAVCSAIIKGNVRCGSCCTNGKCYIDRTG  
LKDNCVFPVDPHONKASTMENQNTNSVFEETENHNOENAPNDONOKNLRSTVEY  
IOSEDFKQITPMTAQPAPFESLQIGRIQVCLVADKSGSLMDRLRNMNASFLY  
LIOTVEGQSWGVAVTFDSAAVYQSELKQNSGADRLILKHLPTVASGTSICSGURT  
AFVIRKKYPTDGEIVLTDGEDNTISSCFDLVQSGAIHFTVALGPAAKELEDS  
KMTGGLQYSSDOVONNGIYDAFAALISGMAIAOHSIOLESNGVNLONNONGSVI  
VDSVSGKDTFLITTTTHTPTTIPIMPSCVCEGFTIDTTTAYALQVPGTAKVPMK  
YSLQASSQITLTVTSRAASATLPITTVFVKNKNGKFPSPYTVASIRQASPILR  
ASVYALIESVNGKTVLELDNAGADATKNDGVSRFTAFDANGRIYSKIMALGCV  
TSDRORAAPKRNAMIDGWIEDGEVRMNPREFTSYVDKQCFSTSGSEFVATN  
VPAAPIPDLFPCCOITDLKASIOGONLVLTWTAGDDYDHGRASNYIIRMSSTSID  
LRHNTGSLQVNTTGLIPKASSEIEFEELGNTGNGTDIPIALQAVDKSKKEI  
SNIRARVTFPAPDBPPEPSTPCPDISTINSITIPCIHVLKIMKVLGEMQVTLGLH"**

**BASE COUNT** 860 a 718 c 693 g 666 t

**ORIGIN**

**Query Match** 9.5%: Score 26; DB 12; Length 2937;

**Best Local Similarity** 100.0%: Pred. No. 0 00023;

**Matches** 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Db** 112 TTCAAAATGTTGCAATTTGATTC 137

234 TTCAAAATGTTGCAATTTGATTC 259

**RESULT** 5

**HU091323** 167880 bp DNA PRI 21-JAN-1999

**LOCUS** Human Chromosome 16 BAC clone CIT987SK-A-972D3, complete sequence.

**DEFINITION** U91323

**ACCESSION** U91323

**NID** 93582311

**VERSION** U91323.1

**KEYWORDS** GI:3582311

**SOURCE** human.

**ORGANISM** Homo sapiens

**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**AUTHORS** 1 (bases 1 to 167880)

Adams, M.D., Loftus, B.J., Phillips, C.A., Zhou, L., Brandon, R. and Venter, J.C.

**TITLE** Direct Submission

**JOURNAL** Submitted (04-MAR-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

**REFERENCE** 2 (bases 1 to 167880)

**AUTHORS** Adams, M.D., Loftus, B.J., Zhou, L., La Bombard, M., Kim, U.J. and Venter, J.C.

**TITLE** Direct Submission

**JOURNAL** Submitted (19-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

**REFERENCE** 3 (bases 1 to 167880)

**AUTHORS** Adams, M.D., Loftus, B.J., Zhou, L., La Bombard, M., Kim, U.J. and Venter, J.C.

**TITLE** Direct Submission

**JOURNAL** Submitted (22-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

**REFERENCE** 4 (bases 1 to 167880)

**AUTHORS** Adams, M.D. and Loftus, B.J.

**TITLE** Direct Submission

**JOURNAL** Submitted (11-SEP-1998) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

**COMMENT** On Sep 11, 1998 this sequence version replaced gi:2335056. BAC clone CIT987SK-972D3 is located in band 16p13.11 of chromosome 16. Genes were identified by a combination of five methods: XGRAL (available by anonymous ftp from artur.epm.ornl.gov), GeneFinder (available by anonymous ftp from collins.washington.edu), GENSCAN (available using the e-mail server at genesan@genomic.stanford.edu), searches of the EST database at TIGR (<http://www.tigr.org/tdb/hcd/hcd.html>) and searches against a peptide database. Repeats were identified using RepeatMasker (Smith, A.F.A. and Green, P. unpublished, <http://ftp.genome.washington.edu/rm/RepeatMasker.html>).

**FEATURES** Location/Qualifiers

**source** 1. 167880

**organism** "Homo sapiens"

**/db\_xref="taxon:9606"**

**/chromosome="16"**

**/map="16p13.11"**

**/clone="A-972D3"**

**/gene="A-972D3.1"**

**/psuedo**

**/gene="A-972D3.1"**

**/gene="A-972D3.2"**

**/gene="A-972D3.2"**

**/gene="A-972D3.2"**

**/gene="A-972D3.2"**

**/gene="A-972D3.2"**

**/product="Myosin heavy chain (MYH11) (3'partial)"**

**/protein\_id="AAC35212.1"**

**/db\_xref="PIR:G1871206"**

**/db\_xref="GI:1871206"**

**/translation="MAQKGLSDDEKFLFVDKFNFSVQAQMAAKRLVWVPSKQG  
FRASTKEKGEVYVVELVNGKRTYVGKDIQKNPKFSEVDEMAELTCLNEASV  
LHNRERFYSGLITY"**

**BASE COUNT** 46429 a 38556 c 38892 g 43951 t

**ORIGIN**

**Query Match** 8.1%: Score 22; DB 11; Length 167880;

**Best Local Similarity** 100.0%: Pred. No. 0.059;

**Matches** 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Db** 32 AAACACATTCACAAATATA 53

25031 AAACACATTCACAAATATA 25052

**RESULT** 6

**HUMRETLAS/c**

**LOCUS** HUMRETLAS 180388 bp DNA PRI 23-NOV-1994

**DEFINITION** Human retinoblastoma susceptibility gene exons 1-27, complete cds.

**ACCESSION** L11910

**NID** 9292420

```
VERSION      L11910.1 GI:292420
KEYWORDS     nuclear protein; recessive oncogene; retinoblastoma gene;
              retinoblastoma protein; retinoblastoma susceptibility; tumor
              suppressor gene.
SOURCE       Homo sapiens DNA.
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Euteiria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (sites)
AUTHORS      Friend,S.H., Bernards,R., Rogel,J.S., Weinberg,R.A., Rapaport,J.M.,
              Albert,D.M. and Dryja,T.P.
TITLE        A human DNA segment with properties of the gene that predisposes to
              retinoblastoma and osteosarcoma
JOURNAL      Nature 323 (6089), 643-646 (1986)
MEDLINE      87039336
REFERENCE    2 (sites)
AUTHORS      Friend,S.H., Horowitz,J.M., Gerber,M.R., Wang,X.F., Bogenmann,E.,
              Li,F.P. and Weinberg,R.A.
TITLE        Deletions of a DNA sequence in retinoblastomas and mesenchymal
              tumors: organization of the sequence and its encoded protein
              [published erratum appears in Proc Natl Acad Sci U S A 1988
              Apr;85(7):2234]
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 84 (24), 9059-9063 (1987)
MEDLINE      88097427
REFERENCE    3 (sites)
AUTHORS      Lee,W.H., Bookstein,R., Hong,F., Young,L.J., Shew,J.Y. and Lee,E.Y.
TITLE        Human retinoblastoma susceptibility gene: cloning, identification,
              and sequence
JOURNAL      Science 235 (4794), 1394-1399 (1987)
MEDLINE      87149066
REFERENCE    4 (sites)
AUTHORS      McGee,T.L., Vandell,D.W. and Dryja,T.P.
TITLE        Structure and partial genomic sequence of the human retinoblastoma
              susceptibility gene
JOURNAL      Gene 80 (1), 119-128 (1989)
MEDLINE      90006771
REFERENCE    5 (sites)
AUTHORS      Sakai,T., Ohtani,N., McGee,T.L., Robbins,P.D. and Dryja,T.P.
TITLE        Oncogenic germ-line mutations in Sp1 and ATF sites in the human
              retinoblastoma gene
JOURNAL      Nature 353 (6339), 83-86 (1991)
MEDLINE      91351319
REFERENCE    6 (bases 1 to 180388)
AUTHORS      Toghiani,J., McGee,T.L., Paterson,J.C., Eagle,J.R., Tucker,S.,
              Vandell,D.W. and Dryja,T.P.
TITLE        Complete genomic sequence of the human retinoblastoma
              susceptibility gene
JOURNAL      Genomics 17 (3), 535-543 (1993)
MEDLINE      94063891
FEATURES     location/Qualifiers
              source          1..180388
              organism="Homo sapiens"
              /db_xref="taxon:9606"
              protein_bind    1858..1863
              /bound_moiety="Spi"
              /evidence-experimental
              protein_bind    1866..1872
              /bound_moiety="ATF"
              /evidence-experimental
              exon            <2060..2196
              /number=1
              CDS             join(2060..2196,5424..5550,39446..39561,41926..42045,
              44668..44706,45799..45866,56853..56963,59651..59793,
              61730..61807,64330..64439,65364..65441,70242..70329,
              73753..73869,76430..76486,76889..76920,77001..77077,
              78083..78279,149998..150116,153208..153353,156693..156838,
              160730..160834,161997..162110,162204..162367,
              170372..170402,173707..173849,174361..174410,
              177005..177078)
              /codon_start=1
              /product="retinoblastoma susceptibility protein"
              /protein_id="AAAS3483.1"
              /db_xref="PID:g292421"
```

```

intron      73870..76429
            /number=13
exon        76430..76486
            /number=14
intron      76487..76888
            /number=14
exon        76889..76920
            /number=15
intron      76921..77000
            /number=15
exon        77001..77077
            /number=16
intron      77078..78082
            /number=16
exon        78083..78279
            /number=17
intron      78280..14997
            /number=17
variation    99426
            /standard_name="Restriction Fragment Length Polymorphism"
            /note="XbaI RFLP"
satellite    102921..102959
            /standard_name="microsatellite RB117"
repeat_region 123912..125501 "Variable Number Tandem Repeats (VNTR)"
            /standard_name="Variable Number Tandem Repeats (VNTR)"
            /rpl_type="tandem"
exon        149998..150116
            /number=18
intron      150117..153207
            /number=18
exon        153208..153353
            /number=19
intron      153354..156692
            /number=19
exon        156693..156838
            /number=20
intron      156839..160729
            /number=20
satellite    156895..156976
            /standard_name="microsatellite RB1.20"
exon        160730..160834
            /number=21
intron      160835..161996
            /number=21
exon        161997..162110
            /number=22
intron      162111..162203
            /number=22
exon        162204..162367
            /number=23

Query Match      7.0%; Score 19; DB 9; Length 180388;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 GACCCAGGCATCTCTGTAT 81
      |||
Db 154951 GACCCAGGCATCTCTGTAT 154933

RESULT 7
LOCUS HS434014 135928 bp DNA PRI 10-JUN-1999
DEFINITION Homo sapiens DNA sequence from PAC 434014 on chromosome 1q32.3.-41.
            Contains the HSD11B1 gene for Hydroxysteroid (11-beta)
            Dehydrogenase 1, the ADORA2BP adenosine A2b receptor like
            pseudogene, the IRF6 gene for Interferon Regulatory Factor 6 and
            two novel genes. Contains ESTs and GSSs, complete sequence.
ACCESSION  AL022398
            G3353547
NID         AL022398.1 GI.3355547
VERSION     HTG: adenosine A2b receptor pseudogene; ADORA2BP; HSD11B1;
            Hydroxysteroid (11-beta) Dehydrogenase 1; Interferon Regulatory
            Keywords

```

SOURCE  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE 1 (bases 1 to 135928)  
JOURNAL  
Graham, D.  
Direct Submission  
Submitted (10-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequests@sanger.ac.uk  
On Jul 30, 1998 this sequence version replaced gi:3046735.  
434014 is from the library RPO13 constructed at the Roswell Park  
Cancer Institute by the group of Pieter de Jong  
For further details see <http://bacpac.med.buffalo.edu/During>  
sequence assembly data is compared from overlapping clones. Where  
differences are found these are annotated as variations together  
with a note of the overlapping clone name. Note that the variation  
annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
EM: EMBL; SW: SWISSPROT; Tr: TREMBL  
This sequence is the entire insert of clone 434014. The true left  
end of clone 447D17 is at 135829 in this sequence. The true right  
end of clone 447D17 is at 2719 in this sequence. This sequence has  
been finished according to sequence map criteria as follows. An  
attempt is made to resolve all sequencing problems, such as  
compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key  
This sequence was generated from a human chromosome 1 bacterial  
clone contig, constructed in collaboration by the Sanger Centre  
chromosome 1 mapping group and Brian Schutte, Bryan Bjork, Kevin  
Coppage and Jeffrey Murray.  
Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr1>.

## FEATURES

## SOURCE

1. 135928

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="q32.3.-41"

/clone="434014"

/clone\_1lb="RPO13"

5. 341

/note="LTRIC repeat: matches 110. 464 of consensus"

500. 557

/note="MER42c repeat: matches 1481. 1538 of consensus"

complement(2439..2713)

/note="match: GSS B14658"

4143. 4145

/note="clone 447D17; CRT in this entry; substitution"

/replace="cgt"

4277. 4279

/note="clone 447D17; CAA in this entry; substitution"

/replace="cga"

5134. 5383

/note="AluSc repeat: matches 50. 299 of consensus;

incomplete repeat"

5382. 5389

/note="clone 447D17; AAAAAAC in this entry; insertion"

/replace="ac"

5405. 5691

/note="AluIo repeat: matches 1. 302 of consensus"

6037. 6483

/note="LTRIC repeat: matches 1. 466 of consensus"

6347. 6349

/note="clone 447D17; GGT in this entry; substitution"

/replace="gct"

6353. 6355

variation



```

/note="clone 447D17; GGG in this entry; substitution"
/replace="gcg"
7546..7628
repeat_region
/note="MIR repeat: matches 84..2 of consensus"
7848..7850
variation
/note="clone 447D17; TTG in this entry; substitution"
/replace="tcg"
8154..8156
variation
/note="clone 447D17; GCA in this entry; substitution"
/replace="gaa"
8161..8448
repeat_region
/note="AluX repeat: matches 1..288 of consensus"
8442..8444
variation
/note="clone 447D17; CGG in this entry; substitution"
/replace="cag"
8449..8478
repeat_region
/note="15 copies of 2 mer 100 % conserved"
8478..8479
variation
/note="clone 447D17; AA in this entry; deletion"
/replace="aacacacacacaca"
8734..9280
repeat_region
/note="LI repeat: matches 3646..4196 of consensus"
10171..10173
variation
/note="clone 447D17; CGT in this entry; substitution"
/replace="cat"
10798..11090
repeat_region
/note="AluX repeat: matches 301..1 of consensus"
11110..11112
variation
/note="clone 447D17; TAC in this entry; substitution"
/replace="tcc"
11472..11474
variation
/note="clone 447D17; GGT in this entry; substitution"
/replace="gct"
join(<11520..11663,13388..14005)
/gene="HSD11B1"
/note="exons 1-4 beyond this clone; match: cDNAs X69561
J05107 S75207 X83203; match: ESTs NT5182 AD121470; match:
genomic DNAs M76663 S63400 M7665 S63400"
/evidence="not_experimental"
/product="d3434014.1 (Hydroxysteroid (11-beta)
Dehydrogenase 1 (EC 1.1.1.146))"
11520..14005
gene
/gene="HSD11B1"
join(<11520..11663,13388..13605)
/gene="HSD11B1"
/note="exons 1-4 beyond this clone; match: proteins P28845
P50172 P51975 Q29608"
P50172
/codon.start=3
/evidence="not_experimental"
/product="d3434014.1 (Hydroxysteroid (11-beta)
Dehydrogenase 1 (EC 1.1.1.146))"
/protein_id="CAA18541.1"
/db_xref="PID:g3112939"
/db_xref="PID:g3355548"
/db_xref="GI:3355548"
/translation="KVAIRPVAAYSASKFALDFFSSIRKYSRVNVSITLCVGL
IDETAKAVSGIVHMOAARKECALEITIGALROEYVDSLWTLTILINPCRI
LEFLYSTYNDPFRINK"
11693..11939
repeat_region
/note="LIPAS repeat: matches 891..641 of consensus"
12266..12741
variation
/note="MER42B repeat: matches 793..1283 of consensus"
12413..12415
variation
/gene="HSD11B1"
/note="clone 447D17; GAA in this entry; substitution"
/replace="gga"
12557..12559
variation
/gene="HSD11B1"
/note="clone 447D17; ATA in this entry; substitution"
/replace="aga"
14614..14616
variation
/note="clone 447D17; ATC in this entry; substitution"
/replace="acc"

```

```

repeat_region 14736..15026
/note="AluSC repeat: matches 2..292 of consensus"
16717..16750
repeat_region
/note="17 copies of 2 mer 82 % conserved"
16745..16748
variation
/note="clone 447D17; GTGT in this entry; insertion"
/replace="gt"
16753..17060
repeat_region
/note="AluO repeat: matches 300..1 of consensus"
17303..17521
repeat_region
/note="LIMD2 repeat: matches 459..673 of consensus"
17472..17474
variation
/note="clone 447D17; TGT in this entry; substitution"
/replace="tat"
17684..18138
repeat_region
/note="MLTIC repeat: matches 1..464 of consensus"
17726..17728
variation
/note="clone 447D17; CTC in this entry; substitution"
/replace="cgc"
17838..17840
variation
/note="clone 447D17; CAT in this entry; substitution"
/replace="cgt"
18278..18280
variation
/note="clone 447D17; ACC in this entry; substitution"
/replace="atc"
18662..18743
repeat_region
/note="LIPB3 repeat: matches 70..150 of consensus"
18744..19043
repeat_region
/note="AluX repeat: matches 2..302 of consensus"
18781..18783
variation
/note="clone 447D17; TCG in this entry; substitution"
/replace="tgc"
19009..19011
variation
/note="clone 447D17; ATG in this entry; substitution"
/replace="acg"
19044..19830
repeat_region
/note="LIPB1 repeat: matches 132..902 of consensus"
19444..19451
variation
/note="clone 447D17; TACACACA in this entry; insertion"
/replace="ta"
19911..20210
repeat_region
/note="Aluub repeat: matches 302..1 of consensus"
20147..20149
variation

```

Query Match 7.0%; Score 19; DB 10; Length 135928;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 AACACTATTCACCAATA 51  
 Db 81633 AACACTATTCACCAATA 81651

```

RESULT 8
LOCUS G48840 320 bp DNA STS 26-MAR-1999
DEFINITION SHGC-83613 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION G48840
NID 94529500
VERSION G48840.1 GI:4529500
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 320)
AUTHORS Myers,R.M.
TITLE Human STS (1999)
JOURNAL Unpublished (1999)
COMMENT
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine

```



```

* 54952 57135: contig of 2184 bp in length
* 57136 57263: gap of unknown length
* 57168 59168: contig of 1905 bp in length
* 59296 59296: gap of unknown length
* 59297 61788: contig of 2492 bp in length
* 61789 65782: gap of unknown length
* 65783 65911: gap of unknown length
* 65911 68324: contig of 2414 bp in length
* 68325 71054: gap of unknown length
* 71055 71182: gap of unknown length
* 71183 73489: contig of 2307 bp in length
* 73490 73617: gap of unknown length
* 73618 76516: contig of 2899 bp in length
* 76517 76645: gap of unknown length
* 76645 81029: contig of 4385 bp in length
* 81030 81157: gap of unknown length
* 81158 87299: contig of 6142 bp in length
* 87300 87427: gap of unknown length
* 87428 95665: contig of 8238 bp in length
* 95666 95793: gap of unknown length
* 95794 104477: contig of 8684 bp in length
* 104478 104605: gap of unknown length
* 104606 119197: contig of 14592 bp in length
* 119198 119325: gap of unknown length
* 119326 125177: contig of 5852 bp in length
* 125178 125305: gap of unknown length
* 125306 138807: contig of 13502 bp in length
* 138808 138934: gap of unknown length
* 138935 149110: contig of 10176 bp in length
* 149111 149237: gap of unknown length
* 149238 164669: contig of 15432 bp in length
* 164670 190000: gap of unknown length
* 164797 190000: contig of 25204 bp in length.

```

```

FEATURES
SOURCE      1. 190000
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="4"

```

```

BASE COUNT  45019 a 47468 c 47837 g 43356 t 6320 others
ORIGIN

```

```

Query Match      7.0%; Score 19; DB 34; Length 190000;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      33 AACACTCATTCACAATA 51
         |||||
Db 149929 AACACTCATTCACAATA 149911

```

```

RESULT 10
LOCUS      AC006874 306131 bp DNA HTG 24-FEB-1999
DEFINITION Caenorhabditis elegans clone Y54E10, WORKING DRAFT SEQUENCE, 3
ACCESSION  AC006874
NID         94263474
VERSION     AC006874.1 GI:4263474
KEYWORDS    HTG; HTGS_P1HASE1.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans; Nematoda; Secernentea; Rhabditia; Rhabditida;
             Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 306131)
            Waterston,R.H.
AUTHORS     The sequence of Caenorhabditis elegans clone
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 306131)
            Waterston,R.H.
AUTHORS     Direct Submission
TITLE       Direct Submission

```

## JOURNAL

Submitted (24-FEB-1999) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 28533: contig of 28533 bp in length
* 28534 28548: gap of unknown length
* 28549 63879: contig of 35331 bp in length
* 63880 63894: gap of unknown length
* 63895 306131: contig of 242237 bp in length.

```

## FEATURES

```

SOURCE      1. 306131
             /organism="Caenorhabditis elegans"
             /db_xref="taxon:6239"
             /clone="Y54E10"

```

```

BASE COUNT  101407 a 53175 c 52480 g 99039 t 30 others
ORIGIN

```

```

Query Match      7.0%; Score 19; DB 34; Length 306131;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      112 TTCAAAATGTCGCAATT 130
         |||||
Db 294459 TTCAAAATGTCGCAATT 294441

```

## RESULT 11

```

CEY105C5.5/c
WPCOMMENT

```

```

Sequence split into 8 fragments LOCUS CEY105C5 Accession 298855
Fragment Name      Begin      End
CEY105C5_0         1      110000
CEY105C5_1        100001    210000
CEY105C5_2        200001    310000
CEY105C5_3        300001    410000
CEY105C5_4        400001    510000
CEY105C5_5        500001    610000
CEY105C5_6        600001    710000
CEY105C5_7        700001    730966
Continuation (6 of 8) of CEY105C5 from base 500001 (298855 Caenorhabditis elegans chr

```

```

Query Match      7.0%; Score 19; DB 34; Length 110000;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      105 ATTATTTCAAAATGTT 123
         |||||
Db 28899 ATTATTTCAAAATGTT 28881

```

```

RESULT 12
LOCUS      AC007115 180821 bp DNA HTG 27-MAY-1999
DEFINITION Homo sapiens chromosome 12 clone 91705, WORKING DRAFT SEQUENCE, in
ACCESSION  AC007115
NID         94895146
VERSION     AC007115.1 GI:4895146
KEYWORDS    HTG; HTGS_P1HASE2.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 180821)

```

AUTHORS Montgomery,K.T., Lau,S.T. and Kucherlapati,R.  
TITLE High Throughput Sequencing of Human Chromosome 12  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 180821)  
AUTHORS Montgomery,K.T., Lau,S.T. and Kucherlapati,R.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAY-1999) Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461, USA  
COMMENT \* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
source 1..180821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="91705"  
/clone\_1id="RPC111"  
BASE COUNT 59926 a 31181 c 31322 g 58392 t  
ORIGIN  
Query Match 7.0%; Score 19; DB 35; Length 180821;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 19 GTCCAGAGATGAACAC 37  
Db 22328 GTCCAGAGATGAACAC 22346  
RESULT 13  
AC007593/c  
LOCUS AC007593 55629 bp DNA HTG 24-MAY-1999  
DEFINITION Drosophila melanogaster chromosome 2 clone BACR01C10 (Db20) RPCI-98  
01.C.10 map 42E-43A strain Y2; cn bw sp, WORKING DRAFT SEQUENCE, 29  
unordered pieces.  
AC007593  
NID 9485670  
VERSION AC007593.3 GI:4885670  
KEYWORDS HTG; HTGS; PHASE1.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 55629)  
AUTHORS Gelniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazek,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chen,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomolan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Segueira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
TITLE Sequencing of Drosophila melanogaster  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 55629)  
AUTHORS Gelniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazek,R.G.,  
Butenhoff,C., Champe,M., Chavez,C., Chen,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomolan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Segueira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
COMMENT On May 24, 1999 this sequence version replaced g1:4883580.

For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 400 bases, phrap computed error  
rate <= 1/10  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 29 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
739 738: contig of 738 bp in length  
819 818: gap of unknown length  
1565 1565: contig of 747 bp in length  
1645 1645: gap of unknown length  
2488 2488: contig of 843 bp in length  
2588 2588: gap of unknown length  
3255 3255: contig of 687 bp in length  
3355 3355: gap of unknown length  
3394 3394: contig of 559 bp in length  
3895 3895: gap of unknown length  
3974 3974: contig of 1099 bp in length  
5073 5073: gap of unknown length  
5074 5074: contig of 1128 bp in length  
5154 5154: gap of unknown length  
6282 6282: gap of unknown length  
6362 6362: contig of 913 bp in length  
7275 7275: gap of unknown length  
7355 7355: contig of 1576 bp in length  
8931 8931: gap of unknown length  
9010 9010: gap of 914 bp in length  
9924 9924: contig of 1254 bp in length  
10005 10005: gap of unknown length  
11258 11258: contig of 1254 bp in length  
11338 11338: gap of unknown length  
12185 12185: contig of 847 bp in length  
12186 12186: gap of unknown length  
12266 12266: contig of 1217 bp in length  
13483 13483: gap of unknown length  
13563 13563: gap of unknown length  
14741 14741: contig of 1179 bp in length  
14821 14821: gap of unknown length  
15969 15969: contig of 1148 bp in length  
15970 15970: gap of unknown length  
16049 16049: gap of 1687 bp in length  
17736 17736: contig of 1687 bp in length  
17737 17737: gap of unknown length  
17816 17816: gap of 1076 bp in length  
18892 18892: contig of 1076 bp in length  
18933 18933: gap of unknown length  
20387 20387: contig of 1415 bp in length  
20388 20388: gap of unknown length  
20467 20467: gap of 1312 bp in length  
21780 21780: gap of unknown length  
21859 21859: gap of 1332 bp in length  
23392 23392: contig of 1332 bp in length  
23471 23471: gap of unknown length  
25449 25449: contig of 1978 bp in length  
25529 25529: gap of unknown length  
25530 25530: gap of 3567 bp in length  
29097 29097: gap of unknown length  
29176 29176: gap of 2049 bp in length  
29177 29177: contig of 2049 bp in length  
31226 31226: gap of unknown length  
31266 31266: gap of 2929 bp in length  
34235 34235: gap of unknown length  
34315 34315: gap of 2850 bp in length  
37165 37165: gap of unknown length  
37244 37244: contig of 4010 bp in length  
37245 37245: gap of unknown length  
41255 41255: contig of 4006 bp in length  
45340 45340: gap of unknown length  
45420 45420: gap of 4531 bp in length  
45421 45421: contig of 4531 bp in length  
49952 49952: gap of unknown length  
50031 50031: gap of 55629 bp in length  
55629 55629: contig of 5598 bp in length  
1..55629 Location/Qualifiers  
FEATURES  
source

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR01C10 (D620) RPCI-98 01.C.10"
/chromosome="2"
/clone.lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC3.6"
/map="42E-43A"
BASE COUNT      14037 a 12352 c 12473 g 14526 t 2241 others
ORIGIN

Query Match      7.0%; Score 19; DB 35; Length 55629;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      29 ATGAACACTCATTCAACA 47
|||||
Db 28872 ATGAACACTCATTCAACA 28854

RESULT 14
AC007624      163362 bp      DNA      HTG      25-MAY-1999
LOCUS      Drosophila melanogaster chromosome 2 clone BACR10F15 (D621) RPCI-98
DEFINITION      10.F.15 map 42E-43A strain Y2; cn bw sp, WORKING DRAFT SEQUENCE, 20
unordered pieces.
ACCESSION      AC007624
NID      94887256
KEYWORDS      AC007624.3 GI:4887256
SOURCE      HTG; HTGS_PHASE1.
ORGANISM      Fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 163362)
REFERENCE      1 Celisner,S.E., Agdayani,A., Arcalata,T.T., Baxter,E., Blazej,R.G.,
AUTHORS      Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacelb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Sait,E.,
Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 163362)
REFERENCE      1 Celisner,S.E., Agdayani,A., Arcalata,T.T., Baxter,E., Blazej,R.G.,
AUTHORS      Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacelb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Sait,E.,
Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (21-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On May 25, 1999 this sequence version replaced g1:4885669.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgrefr@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 400 bases, phrap computed error
rate <= 1/10.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

```

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1      846: contig of 846 bp in length
*      926: gap of unknown length
*      927      1944: contig of 1018 bp in length
*      1945      2024: gap of unknown length
*      2025      3619: contig of 1595 bp in length
*      3620      3699: gap of unknown length
*      3700      5574: contig of 1875 bp in length
*      5575      5654: gap of unknown length
*      5655      7129: contig of 1475 bp in length
*      7130      7209: gap of unknown length
*      7210      9028: contig of 1819 bp in length
*      9029      9108: gap of unknown length
*      9109      11392: contig of 2284 bp in length
*      11393      11472: gap of unknown length
*      11473      15071: contig of 3599 bp in length
*      15072      15151: gap of unknown length
*      15152      17100: contig of 1949 bp in length
*      17101      17180: gap of unknown length
*      17181      20327: contig of 3147 bp in length
*      20328      20407: gap of unknown length
*      20408      26429: contig of 6022 bp in length
*      26430      26509: gap of unknown length
*      26510      32656: contig of 6147 bp in length
*      32657      32736: gap of unknown length
*      32737      38186: contig of 5450 bp in length
*      38187      38266: gap of unknown length
*      38267      47759: contig of 9493 bp in length
*      47760      47839: gap of unknown length
*      47840      54358: contig of 6519 bp in length
*      54359      54438: gap of unknown length
*      54439      62971: contig of 8533 bp in length
*      62972      63051: gap of unknown length
*      63052      76828: contig of 13777 bp in length
*      76829      76908: gap of unknown length
*      76909      104731: contig of 27823 bp in length
*      104732      104811: gap of unknown length
*      104812      128109: contig of 23298 bp in length
*      128110      128189: gap of unknown length
*      128190      163362: contig of 35173 bp in length.
*      Location/Qualifiers
*      1.163362
*      /organism="Drosophila melanogaster"
*      /strain="Y2; cn bw sp"
*      /db_xref="taxon:7227"
*      /clone="BACR10F15 (D621) RPCI-98 10.F.15"
*      /chromosome="2"
*      /clone.lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC3.6"
/map="42E-43A"
BASE COUNT      44309 a 36102 c 35977 g 45454 t 1520 others
ORIGIN

Query Match      7.0%; Score 19; DB 35; Length 163362;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      29 ATGAACACTCATTCAACA 47
|||||
Db 92274 ATGAACACTCATTCAACA 92292

RESULT 15
CELFI0C1      40014 bp      DNA      INV      02-APR-1996
LOCUS      Caenorhabditis elegans cosmid F10C1.
DEFINITION      U49831
ACCESSION      U49831
NID      91203946

```

VERSION U49831.1 GI:1203946  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Rhabdittidae: Rhabdittidae: Pelodermidae: Caenorhabdittis.  
 Rhabdittis: Rhabdittis: Pelodermidae: Caenorhabdittis.  
 1 (bases 1 to 40014)  
 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A., Saunders, D., Showkhen, R., Smalton, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohlman, P.  
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans  
 TITLE  
 JOURNAL  
 MEDLINE  
 94150718  
 2 (bases 1 to 40014)  
 Hallsworth, K.  
 The sequence of C. elegans cosmid F10C1  
 JOURNAL  
 Unpublished (1996)  
 3 (bases 1 to 40014)  
 Waterston, R.  
 TITLE  
 JOURNAL  
 Submitted (20-Feb-1996) Robert Waterston  
 COMMENT  
 Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University,  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RQ, England  
 e-mail: twenemac@wustl.edu and jessesanger.ac.uk  
 NEIGHBORING COSMID INFORMATION:  
 The 5' cosmid is C18A3, 199 bp overlap; 3' cosmid is K05F1, 200 bp overlap. Actual start of this cosmid is at base position 196 of CEF10C1; actual end is at 6257 of CEF105F1  
 NOTES:  
 Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).  
 FEATURES  
 source  
 Location/Qualifiers  
 1.40014  
 /organism="Caenorhabdittis elegans"  
 /db\_xref="taxon:6239"  
 /map="11"  
 6568..12916  
 /gene="F10C1.7b"  
 6568..10346  
 /gene="f10c1.7a"  
 join(6568..6631,7610..7803,8090..8211,8257..8616,8670..9089,9185..9235,9823..9970,10024..10131,10182..10346)  
 /note="coded for by C. elegans CDNA CEMS081F: coded for by C. elegans CDNA cm0198: coded for by C. elegans CDNA cm06d5: coded for by C. elegans CDNA cm14910: coded for by C. elegans CDNA yk3d10.3: coded for by C. elegans CDNA yk494.5: coded for by C. elegans CDNA yk494.3: similar to intermediate filament protein; alternative splicing to F10C1.7b"  
 /codon\_start=1  
 /protein\_id="AA93405.1"  
 /db\_xref="PID:g1203948"  
 /db\_xref="GI:1203948"  
 /translation="MSAVSYSMHRTTTTSSSHGGVSGAGHAAEEFVSAERKQEQ ONSRLVYSRVOLEDNRKELVIEDLRGSLNDIGQIKFNDISLVKREISE AHSGTIGVEYKVDRLDNDYHRYEARREVERREKTTMGASISOQAELDTNKSRY

AAITDEKRLVADQDLYLQAAKDELDAIVDRRLQAEEDDLKIELEFLGRHSQ  
 EITELRTLLAOPADPDRREFKNEIATLAIKAEYDKITQITQETVETQSIASVE  
 SSIVSKNEAAVFRQEEIRKMSITLIRAKLSLEANSLEERANTLQIOLEDORA  
 YSEELHRDNALFRMEDCOTLLAELOALNTQOTDLTEIATIRKIVSEEGRTYVG  
 OGYYVAAOETTRVPEODDMSGEVOTSSFRHAKGNVSIYECPOGKYIILENTS  
 GSVAIEDSNFEITRVIDQVAFYFRPLSHVITQOCHLTKYGNNSGINSPPSIYME  
 SHPSWGGGQGVETFLNSHGIEKASHIQTVASSR  
 join(6568..6631,7610..7803,8090..8211,8257..8616,8670..9089,9185..9235,12375..12522,12576..12683,12734..12916)  
 /gene="F10C1.7b"  
 /note="coded for by C. elegans CDNA CEMS081F: coded for by C. elegans CDNA cm0198: coded for by C. elegans CDNA cm06d5: coded for by C. elegans CDNA cm14910: coded for by C. elegans CDNA yk3d10.3: coded for by C. elegans CDNA yk494.5: coded for by C. elegans CDNA yk494.3: similar to intermediate filament protein; alternative splicing to F10C1.7"  
 /codon\_start=1  
 /protein\_id="AA93404.1"  
 /db\_xref="PID:g1203947"  
 /db\_xref="GI:1203947"  
 /translation="MSAVSYSMHRTTTTSSSHGGVSGAGHAAEEFVSAERKQEQ ONSRLVYSRVOLEDNRKELVIEDLRGSLNDIGQIKFNDISLVKREISE AHSGTIGVEYKVDRLDNDYHRYEARREVERREKTTMGASISOQAELDTNKSRY AAITDEKRLVADQDLYLQAAKDELDAIVDRRLQAEEDDLKIELEFLGRHSQ EITELRTLLAOPADPDRREFKNEIATLAIKAEYDKITQITQETVETQSIASVE SSIVSKNEAAVFRQEEIRKMSITLIRAKLSLEANSLEERANTLQIOLEDORA YSEELHRDNALFRMEDCOTLLAELOALNTQOTDLTEIATIRKIVSEEGRTYVG OGYYVAAOETTRVPEODDMSGEVOTSSFRHAKGNVSIYECPOGKYIILENTS GSVAIEDSNFEITRVIDQVAFYFRPLSHVITQOCHLTKYGNNSGINSPPSIYME SHPSWGGGQGVETFLNSHGIEKASHIQTVASSR  
 complement(13384..14789)  
 /gene="F10C1.3"  
 complement(join(13384..13503,14616..14789))  
 /gene="F10C1.3"  
 /codon\_start=1  
 /protein\_id="AA93406.1"  
 /db\_xref="PID:g1203949"  
 /db\_xref="GI:1203949"  
 /translation="MLAYLVFDFSNVNLRIHALVTCIOCLTISEHSGLAIEDRDT GTKRSTFRVPSDPAVLVPAVAIFCGKGGTCKGKDSYEDLAADKK"  
 22010..28059  
 /gene="F10C1.2b"  
 join(22010..22142,25467..25717,26015..26136,26182..26970,27081..27282,27330..27437,27895..28059)  
 /gene="F10C1.2b"  
 /note="coded for by C. elegans CDNA CEMS081F: coded for by C. elegans CDNA CEMS081F: coded for by C. elegans CDNA yk366.3: coded for by C. elegans CDNA yk12e7.3: coded for by C. elegans CDNA yk12e7.5: coded for by C. elegans CDNA yk37c3.5: similar to intermediate filament protein (alternative splicing to F10C1.2)"  
 /codon\_start=1  
 /protein\_id="AA93407.1"  
 /db\_xref="PID:g1203950"  
 /db\_xref="GI:1203950"  
 /translation="MSHKSSESEYEMQRTSTQPRVAVSROSQNVYSGNAGSG GRLVAMVMSGSAVGGISPAISANAAASFLEADKEKTLQGLNDRGANTIDYVKL DEONRKLVADDELRGKMGKDTSEIKITYSLSSTARDDIDDAARRAEVDKVARRR DLAEVYRSDIDQOORSDREKISOMTNAJADQSEVEMARFQOLDEKERTVAD NSRIWEELOKARSDLDDETTIRIDFONOVOTLMELEFLRSHVEQVEALQALDAP ADTRFREFNEALAIRDKIDYDILIAAGQKDMESWYKIKYSEOGSANRANMESTO RDEVKARDNDIGDLRGKIGDENKNSLLEKRYOUNLTQITDQOYEALNDRPATIR RMRECOQTLVAELQALDTFKOMLDAEIALIYKMLEGETRVLQYMEQVAKTSLQO QENTDSTSVRGEVSTKTTFORSAGKNTISECPNGEFTGLTSHRKNRDNVEHETI RRLKIDGREIYISIPANVIRKGNLITLIYADOGGINNPESLVDGENTWIGIANVY TSIVNKGDEERATHQKTIQSGQ"  
 25329..28059  
 /gene="F10C1.2a"  
 join(25329..25366,25467..25717,26015..26136,26182..26970,27081..27282,27330..27437,27895..28059)

/gene="F10C1.2a"  
/note="coded for by C. elegans cDNA CESQ81F; coded for by  
C. elegans cDNA CESA13F; coded for by C. elegans cDNA  
yk13c11.3; coded for by C. elegans cDNA yk396.3; coded for  
by C. elegans cDNA yk13c11.5; coded for by C. elegans cDNA  
yk37c3.5; similar to intermediate filament protein"  
/codon\_start=1  
/protein\_id="AA93408.1"  
/db\_xref="PID:g1203951"  
/db\_xref="GI:1203951"  
/translation="MEESAPOAPMSGRVLYKMTMGSAIVGIGISPALSANAKSFL  
EATDEKRTILOGINDLGNVYIDRVKLEEDONRKLVAIDLGRKMGKDTSEIKYSE  
SISTAKRIDDAARRAEVDVYARRHDDLAEYRSRYEDIQORRESDRKISQNTAI  
AAQSEVEMLRARFKOLTPDEKRVYTDNSRIPELQKARSDDDETIGRIDFQNOYOT  
LMELEFLRVHEQVEKELQALIAQAPADTREFKNELALAIRDIDEDTIAKQKQ  
DMESEWYKLVSEVOGSANRANMESTYQRDEVKRMRDNIIDLRGKIGDLLENKSLLEKE  
YONLNYQLTDQROYEALNDRDATALRRMEECQTLVAELQALLDTKOMLDAEIAIYR  
KLEGETRVGLTQWEQAVKTHSLQOENTDSTRSVGEVSTKTFORSAGKNVTIS  
ECPDNGKFKIENSHRNKDENGEHRKRIKDRREIYVSIIPANVYIKGKNLTITAR  
DOGGINNPESELVDEENTWGIQANVYISLVNKGEEERATHQKTIQSGQ"  
complement(28559..30536)  
/gene="F10C1.5"  
complement(join(28559..28773,28831..29052,29467..29559,  
29871..30035,30398..30536))  
/gene="F10C1.5"  
/note="similar to D. melanogaster doublesex protein  
(SP:DSXW\_DROME, P23023)"  
/codon\_start=1  
/evidence="not\_experimental"  
/protein\_id="AA93409.1"  
/db\_xref="PID:g1203952"  
/db\_xref="GI:1203952"  
/translation="MPVSPSSVSVLPPTSLLEQLIRIARERNQTPRCARCRRHGTTSA  
LGHKRYCOMKDCMCAKCTILIERQRYMAAOVALRQOQSEEDARDLEVLIGSAGNA  
NLLDLRLDAGEHGHKNASPTNNNNNTTEKYEEDSGQRSSSPGSEIVYSTS  
PELSESTPNTSGSVIPTSPYKPNQYSMPENFVYGPFRFPMPHFCTPSMS  
DFASASASLAAPAFSAOPTIATSSPSTIFPQAPLDYSHRLMKNEAEVEYRENDKT  
F"  
complement(29306..29389)  
/gene="F10C1.5"  
/note="Leu; codon recognized: TTA"  
30597..30788  
/gene="F10C1.1"  
30597..30788  
/gene="F10C1.1"  
/codon\_start=1  
/evidence="not\_experimental"  
/protein\_id="AA93410.1"  
/db\_xref="PID:g1203953"  
/db\_xref="GI:1203953"

Query Match 7.0%; Score 19; DB 36; Length 40014;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 CGACCCCAATGTGCCAGAA 27  
|||||  
Db 37197 CGACCCCAATGTGCCAGAA 37179

Search completed: August 6, 1999, 11:33:17  
Job time: 6083 sec

**This Page Blank (uspto)**



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☒ OTHER: Upside down Duplex Pages

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**This Page Blank (uspto)**